Sudan University of Science and Technology
Faculty of Computer Science and Information Technology

Ambient Intelligence Assisted Healthcare Monitoring

المحيط الذكى المساعد لرصد الرعاية الصحية

A Dissertation Submitted to the PhD Degree

Presented in Fulfillment of the Requirements of Degree of Doctor of Philosophy in Computer Science

PhD Candidate

Abdelhamid Salih Mohamed Salih

Supervisor

Professor: Ajith Abraham

Jun 2015

Dedication

To soul of my parents, to my dear wife, to my dear daughter Amna, to my dear brothers and sisters.

ACKNOWLEDGEMENT

Above all, I thank Almighty God who help me to complete this dissertation and who surrounded me with so many wonderful people.

I would like to express my gratitude and deepest appreciation to my dear supervisor, **Professor Dr. Ajith Abraham** for this kind supports, sharing of knowledge and experience, continuous encouragement, guidance, feedback, suggestions and ideas towards my research direction.

I also would like to thank the Sudan University of Science and Technology to support my research, mainly the Faculty of Computer Science and Information Technology members special thanks goes to **Prof Dr. Izadin Mohammed Osman** (Advisor for Sudan University director for academic Affairs), **Dr. Mohamed Alhafiz** (The head of Quality and Training department in Faculty of Computer Science and Information Technology), **Dr. Talaat M. Wahby** (Dean of the Faculty of Computer Science and Information Technology)**.**

I would like to acknowledge my committee members for the time they spent in reading and listening to my presentations defense and reviewing my work.

I would like to acknowledge all my Colleagues in University of Science and Technology (Faculty of Science and Technology) for their support and encouragement, special thanks to **Dr.Nour Eldin Abdrahman** (Dean of the Faculty of Computer Science and Technology).

My gratitude also goes to Ministry of Ministry of health in Khartoum state for their support and help to implement part of my research at Khartoum state hospitals. My thanks also to the management and Healthcare monitoring departments at Khartoum state (Sudan) hospitals, for their cooperation in this research.

I am also grateful to my classmates in Phd program by courses special thanks to the Colleagues in group of Prof **Dr. Ajith** research.

ABSTRACT

Monitoring is a process of continuously gathering data and performing real-time analysis, monitoring can improve the assessment of the current state, optimization of the business processes, identification of the critical situation, new opportunities, and assisted in decision support and planning. Traditional healthcare monitoring and services are usually offered within hospitals or medical centers, patient's vital signs measurements is conducted by traditional measurements approach. This is costly, inefficient and inconvenient for the people with the need of routine checks. Thus addressing issues pertinent to healthcare monitoring. This dissertation is focusing on the investigation of Ambient Intelligence (AmI) assisted healthcare monitoring model. The data set was constructed by simulation of patients wearable sensors from the environment of Baraha Medical City in Shambat, Khartoum North, Sudan. In connection, with this, we defined and developed proposed integrated AmI healthcare monitoring architecture framework. This research followed a research scientific design approach employing specific methods, including literature review, qualitative data analysis, and data mining (DM) techniques. Literature review was used to define theoretical background and relate the result to the knowledge in the area. In same line to this, Zachman Framework (ZF) was used to guide the development of the AmI healthcare monitoring Information Architecture (IA). Extensive investigation to develop a new novel ensemble health care decision support system for assisting an intelligent health monitoring system were carried and also focusing to reduce the dimensionality of the attributes was done. Extensive investigations of the experimental results of the performance of different Meta classifiers techniques for classifying the data from different wearable sensors used for monitoring different diseases was carried. Results have shown that the architectural representation guided by the selected framework provide a holistic view to the management of healthcare monitoring data. This IA can serve as a strategic guide to the review and development of the healthcare monitoring data collection and analysis systems. The development of AmI healthcare monitoring IA in an enterprise view in the study and design of a AmI healthcare monitoring IA is original contribution of this research, which improves and expands the conceptual framework of the research in this field. Moreover, in addition to identification of wearable sensors vital signs information requirement, the classification of patients situation through novel ensemble decision support and healthcare monitoring system using advanced data mining methods. Evaluation of the research showed that both the process and result of this research are valid and acceptable. The result of this research, mainly the AmI healthcare monitoring, will help healthcare monitoring organizations to revisit their focus of attention in drafting and implementing measures to reduce healthcare monitoring problems. Finally, result of the research can also be replicated to other developing countries with similar context.

الملخص

الرصد هو عملية جمع البيانات باستمرار والقيام بتحليلها في الوقت الحقيقي . عملية الرصد يمكن أن تحسن تقدير الحالة

الراهنة ، تحسن من أساليب العمل، تحديد الوضع الحرج، توفير فرص جديدة، و مساعدة في دعم القرار والتخطيط. وعادة ما تتم عملية رصد رعاية المرضي و خدمات الرعاية الصحية التقليدية داخل المستشفيات أو المراكز الطبية. ويتم قياس العلامات الحيوية للمريض عن طريق نهج القياسات التقليدية.

هذه الطريقة التقليدية مكلفه، غير فعالة وغير مريحه للمرضي ذوي الحاجة للفحوصات الروتينية. وبالتالي لمعالجة القضايا ذات الصلة برصد الرعاية الصحية. هذه الأطروحة تركز على التحقيق في نموزح المحيط الذكي المساعد لرصد الرعاية الصحية. تم تصميم قاعدة البيانات عن طريق المحاكاة للمرضى بأجهزة استشعار يمكن ارتداؤها و ذلك في بيئة مستشفي البراحة الطبية في شمبات، الخرطوم بحري، السودان.

هذا البحث يتبع نهج التصميم العلمي للأبحاث العلمية بتوظيف أساليب محددة، بما في ذلك مراجعة الأدبيات وتحليل البيانات النوعية و تقنيات تنقيب البيانات. تم إستخدم مراجعة الأدبيات لتحديد الخلفية النظرية وربط النتائج بقاعدة المعرفة في المجال. تم إستخدام نهج البحوث النوعية لتحديد الاحتياجات من المعلومات للمجال رصد الرعاية الصحية وتم استخدام تقنيات التنقيب عن البيانات لتطوير نموذج جديد للمحيط الذكي للرعاية الصحية لدعم و تحديد الاحتياجات من المعلومات وكذلك رصد و تحليل و شرح بيانات المرضى في مجال رصد الرعاية الصحية وهذه مساهمة أصلية لهذا البحث.

في نفس الاتجاه ، تم إستخدام إطار ذكمان كدليل لتطور معمارية معلومات المحيط الذكي لرصد الرعاية الصحية . تجارب تحقق مكثفه قد تم إجراها و تنفيذها بإستخدام تقنيات تنقيب البيانات لتطوير نموذج تجمع فرقة جديد ذكي مساعد لدعم اتخاذ القرارفي مجال الرعاية الصحية و أيضا تم التركز علي تقليل حجم السمات في قاعدة البيانات. أيضا تم التحقيق من النتائج التجريبية لأداء مختلف تقنيات مصنفات ميتا لتصنيف البيانات من مختلف أجهزة الاستشعار قابلة للارتداء المستخدمة لرصد الأمراض المختلفة . وقد تم أجراء التجارب علي قاعدة بيانات محاكاة أجهزة استشعار القابلة للارتداء لقياس العلامات الحيوية في بيئة المستشفى.

وقد أظهرت النتائج أن التمثيل المعماري المسترشد به من ذلك الإطار المحدد يقدم نظرة شمولية لإدارة بيانات مراقبة الرعاية الصحية. هذه المعمارية للمعلومات يمكن أن تكون بمثابة دليل استراتيجي لاستعراض وتطوير نظم جمع البيانات وتحليلها في مجال مراقبة الرعاية الصحية.

وقد تم تطوير معمارية معلومات لمحيط ذكي للمساعدة لرصد الرعاية الصحية من وجهة نظرالمؤسسة . دراسة وتصميم محيط ذكي لرصد الرعاية الصحية تعتبر مساهمة أصليه آخري في هذا البحث، مما يحسن ويوسع الإطار المفاهيمي للبحث في هذا المجال.

بالإضافة إلى تعرف أجهزة استشعار القابلة للارتداء العلامات الحيوية لمتطلبات المعلومات، تم تصنيف بيانات أجهزة الاستشعار للعلامات الحيوية للمرضى من خلال نموذج جديدة (الفرقة) يساعد في دعم اتخاذ القرار في نظام رصد الرعاية الصحية باستخدام أساليب متقدمة في تنقيب البيانات .

و قد أظهرت نتائج تقييم هذا البحث أن كل من هذه العمليات و النتائج في هذا البحث صالحة ومقبولة. وسوف تساعد نتائج هذا البحث، وعلى رأسها المحيط الذكي لرصد الرعاية الصحية المستشفيات لإعادة النظر في أداء نظام رصد الرعاية الصحية و التركيز و الاهتمام في صياغة وتنفيذ تدابير للحد من مشاكل رصد الرعاية الصحية. وأخيرا، نتائج هذا البحث يمكن أيضا أن تعمم إلى بلدان نامية أخرى ذات بيئة مماثلة.

Table of Contents

List of Tables

List of Figures

List of Abbreviation

AmI	Ambient Intelligence
HCM	Health Care Monitoring
CD	Chronic Diseases
WSNs	Wireless sensor networks
RFID	Radio Frequency Identification
EAF	Enterprise Architecture Framework
ML	Machine learning
DM	Data Mining
KDD	Knowledge Discovery in Databases
KDP	The knowledge discovery process
DoDAF	Department of Defense Architecture Framework
FEAF	Federal Enterprise Architecture Framework
TOGAF	Open Group Architectural Framework
ZF	Zachman Framework
EA	Enterprise Architecture
IT	Information Technology.
ANN	Artificial Neural Network
KSOMs	Kohonen Self Organizing Maps
ROC	Receiver Operating Characteristics
AmIHCMIA	Ambient Intelligence healthcare monitoring Information Architure
HCMIA	health care monitoring Information Architure
AUC	Area Under ROC Curve
PART	Partial Decision Trees
LMT	Logistic Model Tree

List of Appendices

List of Authors Publication

Journal Publications

- S.M. Salih and A. Abraham, "A Review of Ambient Intelligence Assisted Healthcare Monitoring," International Journal of Computer Information Systems and Industrial Management, USA Vol. 5, pp. 741-750, 2013.
- S. M. Salih and A. Abraham, "Novel Ensemble Decision Support and Health Care Monitoring System," Journal of Network and Innovative Computing, USA, Vol. 2, pp. 041-051, 2014.
- S.M. Salih and A. Abraham, "Ambient Intelligence Healthcare Monitoring Information Architecture (AIHMIA)," International Journal of Computer Information Systems and Industrial Management, USA, Vol. 7, pp. 041-052, 2015.
- S.M. Salih and A. Abraham, " Computational Intelligence Data Analysis for Decision support and Health Care Monitoring System ", Journal of Network and Innovative Computing, USA, Vol. 3, pp. 088-104, 2015.

Conference Publication

- S. M. Salih and A. Abraham, "Intelligent Decision Support for Real Time Health Care Monitoring System," First International Afro-European Conference for Industrial Advancement, Springer Verag, Germany, pp. 183-192, 2015.

1 INTRODUCTION

There is growing need to supply constant Health Care Monitoring (HCM) and support to patients with Chronic Diseases (CD) especially the disabled, and the elderly. **CD** are becoming the major causes of death .Traditional healthcare and services are usually offered within hospitals or medical centers with traditional monitory of patients with **CD**.

Measurements of vital signs are done with traditional measurements and the corresponding diagnosis are carried out. There is an ever-growing need to supply constant care and support to patients with **CD**, disabled, and elderly. The drive to find more effective ways of providing, such care has become a major challenge for the scientific community .

Ambient Intelligence (AmI) for healthcare monitoring (HCM**)** and personalized healthcare is a promising solution to provide efficient medical services, which could significantly lower down the healthcare cost. **AmI** concept, data mining techniques and Information Architecture (IA), involve describing and modeling how information should be organized in such a way that it serves as a foundation for an informed decision support system.

This research contemplates on the concepts of severity analysis of patient's vital signs monitoring and the development of an enterprise **IA** in **AmIHCM**.

This Chapter introduces the essence of the research background, problem statement and motivation, objectives, and significance of the research. Definition of central concepts to the research work is also presented in the Chapter.

1.1 Background of the Research

It is hard to imagine the management of any healthcare system without monitoring. Monitoring is a process of continuously gathering data and performing real-time analysis. Monitoring can improve the assessment of the current state, optimization of the business processes, identification of the critical situation and new opportunities, and prediction of the future state and planning.

1.1.1 Monitoring Systems

Monitoring systems are information systems aimed at monitoring of various other systems, such as an organization, human body, and living environment. Basic elements of monitoring systems are monitoring processes. For example, there is a system aimed at monitoring and predicting biological, ecological, and physical processes along the Gulf of Mexico coastal lines . In general, such systems are characterized by various data gathering methods such as environment sensors, satellite data, and demographic data, as well as advanced data analysis methods that assess and predict critical situations.

In a hospital **HCM** system, it is necessary to constantly monitor the patient’s vital signs, such as blood pressure (BP) and heart rate (HR) to control their health condition. In traditional **HCM** in hospital vital signs monitoring will be done according to personal situation of patient. Often there are three patients situation in hospital. Patients in primary healthcare, patients in intensive care and patients in hospital rooms, all of them need vital signs monitoring to control their health condition. In this research, we focus on patients in hospital rooms.

1.1.2 Monitoring Processes

Systems monitoring can be decomposed into the following processes:Data gathering: The gathering of data about a monitored system is based on various techniques such as surveys, sensors and satellites, and automatic image analysis

- Data preparation: Data preparation is employed today by data analysts to direct their quality knowledge discovery and to assist in the development of effective and high-performance data analysis application systems. This process has to be done automatically as much as possible . In data preparation process data is prepared for analysis by various techniques such as data extraction, cleaning, joining, and transforming.

- Measuring of performances and events: This process is aimed at the measuring of performances and events related to a monitored system.

- Assessment: This process assesses what the measured values mean for the current and future states of the monitored system. Assessment is not limited to any particular data analysis method; however, it is usually based on advanced data analysis techniques such as assessment models, data mining, and advanced data visualization methods.

- Dissemination of results: This process is focused on providing useful information to those who need it.

- Interpretation of results: In this phase the user interprets the information provided by the monitoring system in order to make appropriate decisions.

The assessment processes have an important role in the proposed methodology for developing healthcare monitoring and assessment. The assessment processes in this dissertation, used advanced data mining analysis techniques.

1.1.3 Ambient Intelligence (AmI)

Researchers have defined **AmI** in different ways. **AmI** proposes new ways of interaction between people and technology, making it suited to the needs of individuals and the environment that surrounds . Wireless sensor networks (WSNs) are used for gathering the information needed by **AmI** environments. The principal device in a **WSN**s is the network node, also called mote. This device, battery powered, has the Radio Frequency Identification (RFID) for the transmission and the reception of the information, an interface between the module and the sensor and a microcontroller. The context is defined as any information used to characterize the situation of an entity, which can be a person, a place or an object .

This information is important for defining the interaction between users and the technology that surround them. For these reasons, it is necessary to continuously keep track of information about the users and their environment. The information may consist of many different parameters such as vital signs (e.g. heart rhythm or blood pressure), etc. Thus, distributed sensors throughout the environment and even the users themselves can collect most of the context information. **WSNs** are used for gathering the information needed by **AmI** environments. Sensor data is collected from disparate sources and later analyzed to produce information that is more accurate, more complete, or more insightful than the individual pieces.

Driving from the concept of **AmI**, we simulated the environment of Baraha Medical City in Shambat, Khartoum North, in Sudan using the framework reported in . The monitoring system was for thirty patients with **CD**.

1.2 Statement of the Problem and Motivation

1.2.1 Motivation

The prime motivating factors for this research include the magnitude of the problem across the world in general, and in developing countries in particular. Patient's vital signs monitoring as a major factor for **HCM** problem needs investigation from various perspectives. The capability and use of **IA** to address the problem of **HCM** from data/information management perspective is another factor, which is also a novel approach.

In addition, the multiple application of data mining (DM) in the analysis and interpretation of wearable sensors monitoring data are worth mentioning. Sensing the absence of a foundational framework in a **AmIHCM** information management in a local context is among the motivations for this research work though there were attempts dealing with an aspect of **AmIHCM** data.

Thus, motivated by dearth of a systematic view in monitoring patients wearable sensors analysis in a **AmI** healthcare monitoring domain and a high-level architectural guide line from integrated enterprise perspective in the **AmIHCM** information management domain in Khartoum State (Sudan) context, this research focuses on identifying determinant patients situation and investigating the potential use of enterprise information architecture (EIA) as an approach to the management of **AmIHCM** data.

1.2.2 Problem Statement

Traditional healthcare and services are usually offered within hospitals or medical centers. Chronic diseases are becoming the major causes of death such as insufficient cardiac heart, asthma, diabetes, and patients with Alzheimer’s disease. In EU countries, the heart disease is the most common cause of death . According to US National Center for Health Statistics, major chronic diseases such as heart disease, cerebrovascular disease, and diabetes account for 35.6% of death in US in the year 2005 . In Sudan according to the latest , data published in April 2011 Coronary Heart Disease Deaths reached 10.67% of total deaths. There is an ever-growing need to supply constant care and support to patients with chronic diseases, disabled, and elderly.

In traditional **HCM**, the patient’s vital signs are measured by experts and the data was recorded in patients vital signs sheet, so as to be presented to doctors for diagnosis purpose and later will be achieved at statistical office. However, the use of these vital signs sheet shows serious limitations, of which the prime cause lies in the instruments and procedures used.

The drive to find more effective ways of providing such care has become a major challenge for the scientific community . Also people in post-surgery state need continuous monitoring of their health condition, especially the vital signs, until their health status becomes stable. Patients, as well as their families, also need to collaborate with their doctor and medical professionals to get informed about their states. Until now, the monitoring of the health condition of such people is usually accomplished within medical centers or hospital environments.

As a result, measurements of vital signs and the corresponding diagnosis are carried out in controlled environments. However, this solution is costly, inefficient and inconvenient for the people with the need of routine checks, since the patients need to frequently visit the hospital, sometimes on a daily basis, or even worse, need a long-stay.

There are huge requirements to move the routine medical check and healthcare services, thus release the hospital beds and other limited resources to the people with urgent needs. **AmI** for **HCM** and personalized healthcare is a promising solution to provide efficient medical services, which could significantly lower down the healthcare cost and improve the health care.

1.3 Research Questions

The central research questions of this research are the following:-

- Does **AmI** provide efficient medical services, which could significantly improve assisted **HCM**? If so, in what way (s)?

There are others sub questions:-

- How a generic **IA** can help in developing the **AmI** assisted health care architecture?
- What are the problems of the current patients vital signs monitoring data collection, analysis and dissemination practice on the healthcare monitoring?
- How knowledge discovered from sensor data could be used in practice?
- How a machine learning designed ensemble based decision support system can optimize the results and improve assisted **HCM** information articture?

1.4 Objective of the Research

1.4.1 General Objective

Our general objective in this research is as follows:-

- To investigate novel ensemble decision support method by intelligent analysis of patient's sensor data using WEKA , to get better results to improve assisted health care monitoring.

1.4.2 Specific Objectives

In addition to the main objective, there are sub objectives as the following:-

- To investigate a way to define information architecture based on enterprise architecture framework (EAF) to establish **AmIHCM** decision support and monitoring information management.

- To define and develop a **AmIHCM** decision support and monitoring **IA** that will facilitate effective utilization and management of **HCM** information among hospital departments.

- To develop a generic Information architecture using Zachman's framework to conceptualize a more generic model.

- To examine problems related to patients vital signs monitoring data reporting, data quality and analysis mechanisms in a **HCM** domain.

- To identify the structure and requirement of **AmIHCM** data collection and analysis focusing on healthcare vital signs monitoring data.

- To simulate the patient sensors data.

- To investigate the experimental results of the performance of different classification techniques for classifying the data from different wearable sensors vital signs used for monitoring different diseases.

- To evaluate the method under several evaluation methods.

1.5 Significance of the Research

HCM is a critical issue in a society. One of the significant factors that affect **HCM** is patients vital signs monitoring; thus, studies on the data collection and analysis of these healthcare patients vital signs monitoring have paramount importance. In this research, an attempt is made to conduct trend analysis, explore **HCM** situations, and develop integrated information architecture in support of **AmIHCM** information management.

In line with this, the major contribution of this research will be explanations of **HCM** situations and the design of an integrated architecture for effective **AmIHCM** information management and utilization among **HCM** organizations to assisted decision support in **HCM**.

The **IA** established in the **AmIHCM** organizations forms the basis of support in delivering future information requirements for the **HCM** domain. It is also worth mentioning the different views and theories that will result from such type of study in different context. Although the study is aimed at addressing **HCM** in particular, the output of the study can be used as a source of methodological approach for studies dealing with the application of **IA** and standards on other societal problems.

In addition to providing an invaluable perspective for **HCM** data collection and analysis, it will subsequently improve the efficiency of a national data collection system. The outcome of the study contributes to the body of knowledge on IA, **AmI HCM**, patient’s vital signs information analysis and management, and might (a) help researchers and **HCM** organizational leaders to base their strategies and plans of **HCM** on a complete and reliable information guided by a wider view of the problem, and (b) provide empirical evidence on different aspects of **HCM** situation, allowing for better understanding of the problem at hand. An additional vital contribution is made by the introduction of the enterprise approach for the study of the healthcare monitoring data management systems.

Thus, results of this research add on the efforts to understand the **HCM** situation in Khartoum State (Sudan) and evaluate the existing **HCM** data system and/or design a new one.

1.6 Scope of the Research

The scope of the research is limited to exploring the existing situation of **HCM** patient’s vital signs monitoring information management practices in hospitals and developing **IA** with the aim of laying down a foundational framework for **AmI HCM** information management. This research would have been more complete if it had included implementation and deplovement of **AmI HCM** application as well. However, for the sake of cost and time, this research is limited to simulate the environment of the hospital to monitoring patients wearable sensors vital signs data. In general, the research will focus only on healthcare patient’s vital signs monitoring information management.

1.7 Definition of Working Concepts

Understanding of the ideas and contributions of this dissertation is grounded in the understanding of the key terms and concepts it operates with. To this end, working definitions and explanations of the key terms used throughout the research are presented below :-

Ambient Intelligence (AmI)

AmI have been defined by Researchers in different ways as given below:

"*AmI proposes new ways of interaction between people and technology, making it suited to the needs of individuals and the environment that surrounds them*" .

"Ambient Intelligence': people living easily in digital environments in which the electronics are sensitive to people's needs, personalized to their requirements, anticipatory of their behavior and responsive to their presence .

"A developing technology that will increasingly make our everyday environment sensitive and responsive to our needs .

Framework

A framework helps people to organize and assess integrated models of their enterprises. This organization helps ensure interoperability of systems and control the cost of developing systems .The two definitions below best illustrate the meaning of the term 'Framework':

"A classification structure for descriptive representation of an object, any object. An object could be an enterprise, an information system, etc…" .

"A systematic taxonomy of concepts and their interrelations" .

Architecture Framework

There are various conceptualizations of architectural frameworks. Some focus at a software level, while others try to view at the organizational level. It is also stated that most of the classical enterprise architecture frameworks presented focus on the software architecture, rather than on the total enterprise architecture . A comprehensive definition of the concept ‘architecture framework’ is presented by The Open Group (2002) and is adopted here as follows:

"An architecture framework is a tool, which can be used for developing a broad range of different architectures. It should describe a method for designing information system in terms of a set of building blocks, and for showing how the building blocks fit together. It should contain a set of tools and provide a common vocabulary. It should also include a list of recommended standards and compliant products that can be used to implement the building blocks” .

Information Architecture (IA)

There are many definitions, views and conceptualizations of IA depending on its application. It is noted in , that there is no widely accepted definition. Two definitions are selected here for their comprehensiveness and the fact that they reflect the context of the concept in this research.

“ IA represents a higher level of abstraction, emphasizing an awareness of systems in terms of how critical subcomponents interacts according to semantic aspects of processes, designs, and metrics” .

“IA is a supporting business processes by using methods, techniques and software to design, control and analyze operational processes involving humans, organizations, applications, documents and other sources of information” .

“It is defined, as a professional practice and field of studies focused on solving the basic problems of accessing and using, the big amounts of information available today.” .

Enterprise Architecture Framework (EAF)

“As a kind of implicit conceptual meta model of the architecture of their IT systems. It describes the architecture of a business and its information technology (IT), and their alignment” .

Zachman Framework of Architecture (ZFA)

"Zachman describes the aim of this framework as an architecture that represents the information systems' artifacts, providing a means of ensuring that standards for creating the information environment exist and they are approximately integrated" .

1.8 Contribution

The main contribution of this research is the definition of generic **IA** to investigate novel ensemble methods of **AmI** (interpretable) and to see which method, would work better in the decision support for assisted health care monitoring.

Others contributions in this thesis are as follows:-

- To provide a framework for the integration of future standards developed for data representation, manipulation, and visualization.
- To simulate patient's wearable sensors data.
- To investigate novel ensemble Decision support modules.
- To help and enhance the research capabilities of Sudan University and the Sudan at large in the area of **AmI** assisted heath care monitoring.
- To open new opportunities for collaboration between Sudan University, the industry and hospitals in the area **HCM**.

1.9 Structure and Organization of the Thesis

The remaining part of the dissertation is organized as follows. Chapter Two presents literature review to the central theme of the research. Chapter Three is dedicated to details related work. Chapter Four presents the research approach and methodology, while Chapter five focus on data analysis and experimentation, investigate simulated wearable sensors monitoring patients data and reduction the wearable sensors data size by removing irrelevant and redundant attributes and in developing novel intelligent ensemble health Care decision support and monitoring system. The chapter presents Data mining and machine learning experimentation and results, qualitative data analysis, findings and conclusion formed from the data analysis and experimentation. Chapter six presents the proposed integrated information architecture for **AmIHCM** domain, the features identified through experiments, insights regarding data quality problems and results of qualitative data analysis are used in developing architectural description under the six dimensions. Chapter seven presents evaluation of the results**.** Issues of validity, reliability and generalizability of findings, accuracy and interestingness of data mining and machine learning experiments, completeness, practical utility, and robustness of the architectural models are examined and explained. Chapter eight presents' conclusions from the whole research process and summary of the results, contributions of the research, and identified future research lines.

2 LITERATURE REVIEW

2.1. Overview

This Chapter deals with the AmI healthcare monitoring concepts, Machine learning (ML)/Data Mining (DM) and **EAF**. Section 2.2 covers the foundation of **AmI** . Section 2.3 includes the discussion of **ML** and **DM** in general. The application and theoretical foundation of **EAF** are illustrated in Section 2.4.

2.2 AmI Healthcare Monitoring

Traditional healthcare and services are usually offered within hospitals or medical centers. **CD** is becoming the major causes of the death. In EU countries, the heart disease is the most common cause of death . According to US National Center for Health Statistics, major **CD** such as heart disease, cerebrovascular disease, and diabetes account for 35.6% of death in US in the year 2005, . In Sudan according to the latest WHO , data published in April 2011, Coronary Heart Disease (CHD) death's reached 10.67% of total deaths. There is an ever-growing need to supply constant care and support to patients with **CD**, disabled, and elderly.

The drive to find more effective ways of providing, such care has become a major challenge for the scientific community, as pointed out by Nealon . Also people in post-surgery state need continuous monitoring of their health condition, especially the vital signs, until their health status becomes stable. Patients, as well as their families, also need to collaborate with their doctor and medical professionals to get informed about their states. Until now, the monitoring of the health condition of such people is usually accomplished within medical centers or hospital environments.

As a result, measurements of vital signs and the corresponding diagnosis are carried out in controlled environments. However, this solution is costly, inefficient and inconvenient for the people with the need of routine checks, since the patients need to frequently visit the hospital, sometimes on a daily basis, or even worse, need a long-stay.

There are huge requirements to move the routine medical check and healthcare services from Traditional healthcare monitoring to a new paradigm . Thus **AmI** for healthcare monitoring and personalized healthcare is a promising solution to provide efficient medical services, which could significantly lower down the healthcare cost and improve the healthcare monitoring.

As stated by Wesiser **AmI** is an emerging multidisciplinary area based on ubiquitous computing, which influences the design of protocols, communications, systems, devices, etc. According to Tapia et al., **AmI** proposes new ways of interaction between people and technology, making it suited to the needs of individuals and the environment that surrounds those. It also proposes a new way to interact between people and technology, where this last one is adapted to individuals and their context. The context includes both the users and the environment information.

The information may consist of many different Parameters such as the building status (e.g. temperature or light), vital signs (e.g. heart rhythm or blood pressure), etc. **WSNs** are used for gathering the information needed by **AmI** environments. Some examples of possible **WSNs** technologies are Radio Frequency Identification (**RFID)**, ZigBee or Bluetooth. Gather information about the context is not enough. However information must be processed, analyzed, reasoning and decision support, since the quality of decision support depends on quality of information by using dynamic mechanisms and methods. In this sense, various Architectures and models have been used for the development of **AmI** systems.

2.2.1 Ambient Intelligence (AmI)

The European Commission's Information Society Technologies Advisory Group (ISTAG), has introduced the concept of AmI. Researchers have defined AmI in different ways as given below. These definitions are summarized, in addition to highlighting the features that are expected in AmI technologies: sensitive, responsive, adaptive, transparent, ubiquitous, and intelligent.

- **AmI** is an emerging multidisciplinary area based on ubiquitous computing which influences the design of protocols, communications, systems, devices, etc. .

- **AmI** proposes new ways of interaction between people and technology, making it suited to the needs of individuals and the environment that surrounds those .

- A developing technology that will increasingly make our everyday environment sensitive and responsive to our needs .

- A potential future in which we will be surrounded by intelligent objects and the environment will recognize the presence of persons and will respond to it in an undetectable manner .

- **AmI** implies intelligence that is all around us (Maeda and Minami, 2006) [29].

- The presence of a digital environment that is sensitive, adaptive, and responsive to the presence of people .

2.2.2 Areas Related with AmI

AmI inherits aspects of many areas of Computer Science (Figure 2.1), but should not be confused with any of those in particular. Networks, Sensors, Human Computer Interfaces (HCI), Pervasive Ubiquitous Computing and Artificial Intelligence (AI) are all relevant and interrelated but none of them conceptually covers the full scope of AmI. AmI puts together all these resources to provide flexible and intelligent services to users acting in their environments .

Ambient Intelligence aims at taking the embedding of devices one step further by involving the entire environment, i.e., any physical object, in the interaction with people, thus integrating electronics fully into the background of people with the purpose of improving productivity, creativity, and pleasure through enhanced user-system interaction. Evidently, the **AmI** vision uses solutions from the earlier visions on Mobile and Pervasive Computing. There are also new elements that call for novel approaches.

These can be best explained from the two key words in the notion of Ambient Intelligence. The word ambience in **AmI** refers to the need for a large-scale embedding of technology in a way that it becomes non-obtrusively integrated into every-day objects and environments. The word intelligence reflects that the digital surroundings exhibit specific forms of social interaction, i.e., the environments should be able to recognize people, personalize to their individual preferences, adapt themselves to users over time, and possibly act upon their behalf.

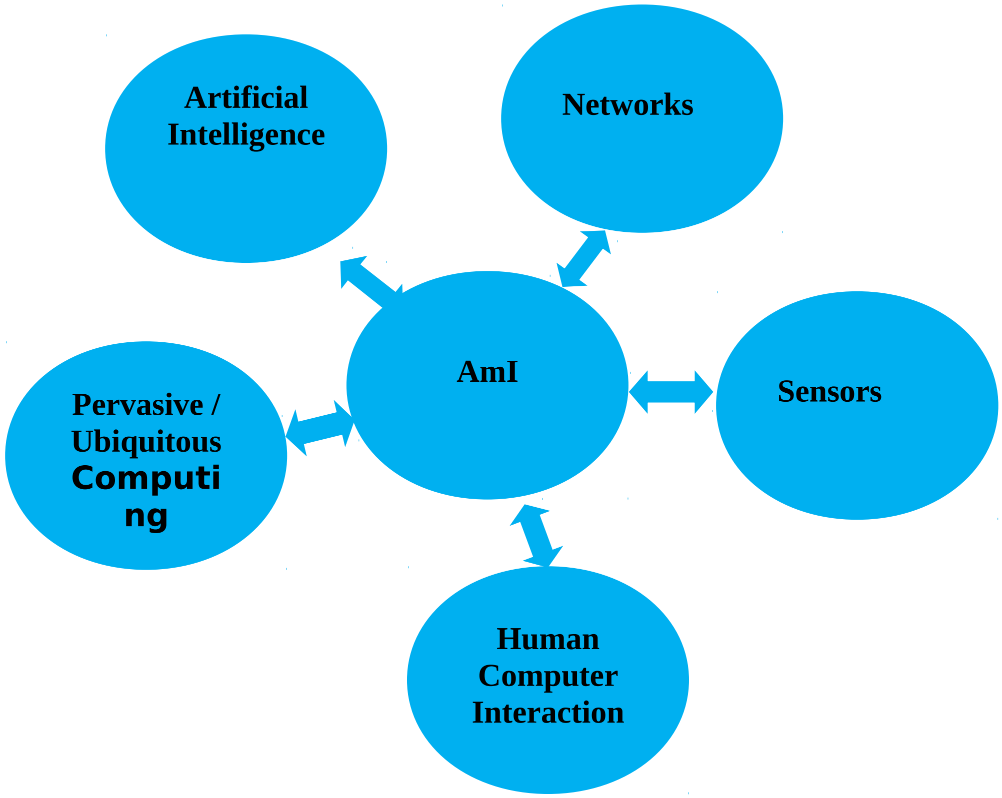


Figure 2.1: Areas related with AmI999

This implies that embedding through miniaturization is the main systems design objective from a hardware point of view. From software point of view the main objective is to introduce true intelligence into these systems.

2.2.3. Technologies Used in AmI

WSNs are used for gathering the information needed by **AmI** environments. **WSNs** research has become a popular area of research in recent years . The **WSNs** community has explored applications such as environmental monitoring, situational awareness, and structural safety monitoring. have described different topologies for different scenarios.

The main topologies used in **WSN** are Star Network and Mesh Network. Some examples of possible **WSNs** technologies are radio **RFID**, ZigBee or Bluetooth. . However ZigBee is a low-cost, low-power, wireless mesh networking standard. The low cost option allows the technology to be widely deployed in wireless control and monitoring applications, the low power usage allows longer life with smaller batteries, and the mesh networking provides high reliability and larger range.

The principal device in a **WSN**s is the network node, also called mote. This device, battery powered, has the **RFID** for the transmission. and the reception of the information, an interface between the module and the sensor and a microcontroller. The context is defined as any information used to characterize the situation of an entity, which can be a person, a place or an object . This information is important for defining the interaction between users and the technology that surround them. For these reasons, it is necessary to continuously keep track of information about the users and their environment.

The information may consist of many different parameters (sensors) such as location, the building status (e.g. temperature or light), vital signs (e.g. heart rhythm or blood pressure), etc. Thus, distributed sensors throughout the environment and even the users themselves can collect most of the context information.

WSNs are used for gathering the information needed by AmI environments, whether in home automation, educational applications or healthcare monitoring . The same can be said about a Body Area Network (BAN) that allows the monitoring of vital sign parameters of a user.

Table 2.1: Sensors used to detect vital signs

Sensors	**Observation**
ECG	Heart rate, heart rate variability (HRT)
EMG	Muscle activities and fatigue
Temperature	Skin temperature, health state (fever)
Respiration	Breathing rate, physical activity
Blood oxygen	Status of the cardiovascular system, heart rate
Blood pressure	Status of the Cardiovascular hypertension.

A combination of both networks is also possible including mobile nodes in the body of a person that can be directly connected with the fixed mesh network in the environment. Moreover, it is necessary for the integration and the interoperability of these networks with the Personal, Local, Metropolitan and Wide area networks used in the “classical” applications.

The same **WSNs** used in the connection of sensors, allow the transmission of commands and data to actuators in the environment (engines, lights, etc.) and user body (vibrator, leds, etc.). Table 1 highlights the sensors used to detect vital signs and Table 2.2 highlights the sensors used to detect motion and location.

Table 2.2: Sensors used to detect motion and location

Sensors	**Observation**
Accelerometer	Motion patterns of the body and limbs
Microphone	Speaker recognition, localization by ambient sound, activity detection, speech features
Visible light sensor	Location of Light sources
Rotation	Body movement
Compass	Orientation of the body and head
Air Pressure	Vertical motion in elevator or staircase
Light sensor	Sunshine, location of lamps
Environment temperature	Outdoor, indoor
Wlan/ GSM/ CDMA	Location, user environment
Bluetooth, ZigBee	Services and devices nearby

2.2.4 Requirements for Wireless Medical Sensors

Wireless medical sensors should satisfy the main requirements such as wearability, reliability, security, and interoperability .

- Wearability: It requires that the wireless medical sensors should be lightweight and small.

- Reliable communication: One approach to improve reliability is to move beyond telemetry by performing on sensor signal processing.

- Security: A relatively small number of nodes in a typical WWBAN and short communication ranges make key establishment, authentication, and data integrity achievable.

- Interoperability: Wireless medical sensors should allow users to easily assemble a robust WWBAN depending on the user's state of health.

2.2.5 Motivation

We can expect a greater demand for services and applications oriented towards people with **CD**. These demands include providing services for those who suffer from various illnesses and the need for constant healthcare monitoring such as: diabetes, arthritis, senile dementia, Alzheimer, heart-related diseases among many others. **AmI** and the emergence of new types of mobile and embedded computing devices, developments in wireless networking, smart sensors, and others, gives us the tools and methods to come up with innovative applications to better assist users, and therefore, improve their lifestyle. We believe that support for people with chronic diseases and elderly through **AmI** infrastructures is the most naturally appropriate. Not just because of its evident social impact, but because of the characteristics of its special requirements. The objective of **AmI** is to develop intelligent and intuitive systems and interfaces, capable of recognizing and responding to the user’s necessities in a ubiquitous way, providing capabilities for ubiquitous computation and communication, considering people in the center of the development, and creating technologically complex environments in medical, domestic, academic and other fields .

2.3 Data Mining (DM) and Machine Learning (ML)

2.3.1 Overview

There is growing need to supply constant **HCM** and support the patients with **CD** especially the disabled, and elderly. Wireless sensor networks (WSNs) are used for gathering the information needed. The information may consist of many different sensors such as vital signs (e.g. heart rhythm or blood pressure), etc. Thus, most of the context information can be collected by distributed sensors throughout the environment and even the users themselves .

When analyzing sensor data, **AmI** systems may employ a centralized or distributed model . Sensors in the centralized model transmit data to a central server, which fuses and analyzes the data it receives. In the distributed model, each sensor has onboard processing capabilities and performs local computation before communicating partial results to other nodes in the sensor network. The choice of model will have a dramatic effect on the computational architecture and type of sensor that is used for the task as described by Benini and Poncino , and also by Jayasimha . In both cases, sensor data is collected from disparate sources and later analyzed to produce information that is more accurate, more complete, or more insightful than the individual pieces.

There are several machine learning techniques and data mining methods and techniques used in analyzing sensors data in **AmI**. These methods and techniques can help accomplish many important tasks in AmI assisted healthcare monitoring and make the system more efficient .

In health monitoring systems, focus has been recently shifting from that of obtaining data to one of developing intelligent algorithms to perform a variety of the tasks . Such tasks not only include traditional pattern recognition and anomaly detection but also must consider decision support systems. These latter challenges are particularly important if health monitoring services are to be designed that can address the growing market needs and long-term monitoring. This Section and the following attempts to clarify how certain data mining methods have been applied in the literature. It also attempts to reveal trends in the selection of the data processing methods based on the requirements of the healthcare monitoring system. Focus is put on the data mining / machine learning algorithms that have been used in order to provide an overview of the algorithm's capabilities and shortcomings. These methods and techniques can help accomplish many important tasks in AmI assisted healthcare monitoring and make the system more efficient.

2.3.2. What is Data Mining / Machine Learning

The objective of data mining is to identify valid, novel, potentially useful, and understandable correlations and patterns in existing data . Finding useful patterns in data is known by different names (including data mining) in different communities (e.g., knowledge extraction, information discovery, information harvesting, data archeology, and data pattern processing) .Various definitions for data mining exist in the literature. defines defined data mining as ‘‘the process that uses statistical, mathematical, artificial intelligence and machine-learning techniques to extract and identify useful information and subsequently gain knowledge from large databases”. In data mining is defined as ‘the task of discovering interesting patterns from large amounts of data where the data can be stored in databases, data warehouses, or other information repositories’. Authors in also provided a similar definition regarding data mining as being the process of extracting or detecting hidden patterns or information from large databases. The term “data mining” is primarily used by statisticians, database researchers, and the MIS and business communities. The term Knowledge Discovery in Databases (KDD) is generally used to refer to the overall process of discovering useful knowledge from data, where data mining is a particular step in this process .

Data mining is an extension of traditional data analysis and statistical approaches in that it incorporates analytical techniques drawn from a range of disciplines including, but not limited to. Literature agrees to classify data mining technology into either predictive or descriptive modeling. The goal of a predictive modeling is to predict the value of one column based on the value of other columns. Descriptive modeling on the other hand deals with describing all of the data and discovering patterns and segments of the data . Thus, it is also called unsupervised technique. The most widely used tasks are classification, association, clustering, dependency analysis, prediction, segmentation and description .

The data-mining component of the **KDD** process is concerned with the algorithmic means by which patterns are extracted and enumerated from data. The process includes the evaluation and possible interpretation of the mined patterns to determine which patterns can be considered as new knowledge.

Before one attempts to extract useful knowledge from data, it is important to understand the overall approach. Simply knowing many algorithms used for data analysis is not sufficient for a successful **DM** project. The process defines a sequence of steps (with eventual feedback loops) that should be followed to discover knowledge (e.g., patterns) in data. Each step is usually realized with the help of available commercial or open-source software tools. The knowledge discovery process (KDP), also called knowledge discovery in databases, seeks new knowledge in some application domain. It consists of many steps (one of them is **DM**), each attempting to complete a particular discovery task and each accomplished by the application of a discovery method.

2.3.3 Knowledge Discovery Process (KDP)

The first basic structure of the model was proposed by , and later improved/modified by others. The process consists of multiple steps that are executed in a sequence. Each subsequent step is initiated upon successful completion of the previous step, and requires the result generated by the previous step as its input. There are others KDP models developed by researchers, but the main differences between the models lie in the number and scope of their specific steps. A common feature of all models is the definition of inputs and outputs. We restrict our discussion to those models that are popular in the literature and have been used in real knowledge discovery projects. The two process models developed in 1996 and 1998 are the nine-step model by Fayyad et al. and the eight-step model by Anand et al., . Below we introduce the first of these, which is perceived as the leading research model.

The model consists of nine steps, which are outlined as follows:

1. Developing and understanding the application domain. This step includes learning the relevant prior knowledge and the goals of the end user of the discovered knowledge.

2. Creating a target data set. Here the user selects a subset of variables (attributes) and data points (examples) that will be used to perform discovery tasks. This step usually includes querying the existing data to select the desired subset.

3. Data cleaning and preprocessing. This step consists of removing outliers, dealing with noise and missing values in the data, and accounting for time sequence information and known changes.

4. Data reduction and projection. This step consists of finding useful attributes by applying dimension reduction and transformation methods, and finding invariant representation of the data.

5. Choosing the data-mining task. Here the user matches the goals defined in Step 1 with a particular **DM** method, such as classification, regression, clustering, etc.

6. Choosing the data-mining algorithm. The user selects methods to search for patterns in the data and decides, which models and parameters of the methods used may be appropriate.

7. Data mining. This step generates patterns in a particular representational form, such as classification rules, decision trees, regression models, trends, etc.

8. Interpreting mined patterns. Here the analyst performs visualization of the extracted patterns and models, and visualization of the data based on the extracted models.

9. Consolidating discovered knowledge. The final step consists of incorporating the discovered knowledge into the performance system, and documenting and reporting it to the interested parties. This step may also include checking and resolving potential conflicts with previously believed knowledge.

This process is iterative. The authors of this model declare that a number of loops between any two steps are usually executed, but they give no specific details. The model provides a detailed technical description with respect to data analysis but lacks a description of business aspects. This model has become a cornerstone of later models. The model has been used in a number of different domains, including engineering, medicine, production, e-business, and software development.

2.3.4 Data Mining Tasks

define six main functions for **DM**:

- Classification is finding models that analyze and classify a data item into several predefined classes.
- Regression: is mapping a data item to a real-valued prediction variable.
- Clustering: is identifying a finite set of categories or clusters to describe the data.
- Dependency Modeling (Association Rule Learning) is finding a model which describes significant dependencies between variables
- Deviation Detection (Anomaly Detection): is discovering the most significant changes in the data.
- Summarization: is finding a compact description for a subset of data

DM has two primary objectives of prediction and description. Prediction involves using some variables in the data sets in order to predict unknown values of other relevant variables. It refers to the process of building a model that will permit the value of one variable to be predicted from the known values of other variables . For example: classification, regression, and anomaly detection and description involve finding human understandable patterns and trends in the data. For example: clustering, association rule learning, and summarization.

2.3.5 Data Mining Tasks for Wearable Sensors data

Recently, the research area of health monitoring systems has shifted from simple reasoning of wearable sensor readings to the higher level of data processing in order to give much more information that is valuable to the end users. Therefore, healthcare services have been focusing on deeper data mining tasks to have deeper knowledge representation. Based on the selected literature, three types of data mining tasks are predominant. These three tasks are: prediction, anomaly detection which may include the subtask of raising alarms, and diagnosis where a decision making process is made to often categorize the data into different groups depending on the diseases.

Each of these tasks is further described in the next chapter. Based on the literature, most monitoring applications which consider home settings or remote monitoring deal predominantly with prediction and anomaly detection whereas the applications in clinical settings are typically focused on diagnosis . This fact is easily explained by the growing desire to have a more preventative approach (prediction) via wearable sensors and to consider the possibility to facilitate independent living in home environments by increasing the sense of security (alarm). Similarly, in clinical settings much more information is available in order to provide diagnosis and assist in decision making .

2.3.6 Data Mining Approach

In healthcare monitoring systems, the role of data analysis is to extract information from the low level sensor data and bridge them to the high level knowledge representation. For this reason, recent healthcare monitoring systems have given more attention to the data processing phase in order to extract more valuable information based on the expert user requirements.

Regardless of the **DM** technique used, the most standard and widely used approach to mining information from wearable sensors is given in Figure 2.2. As depicted in Figure 2.2, the raw sensor data is typically used as a starting point of the **DM** approach. The sensor data is provided for both training data in order to learn the system, extract relevant features, as well as test data for real-world usage designed model.

The main steps of the data mining approach consist of: (1) data preprocessing (2) Attribute selection and (3) Data learning models, considering expert knowledge and metadata to perform the tasks such as detection, prediction, and decision support.

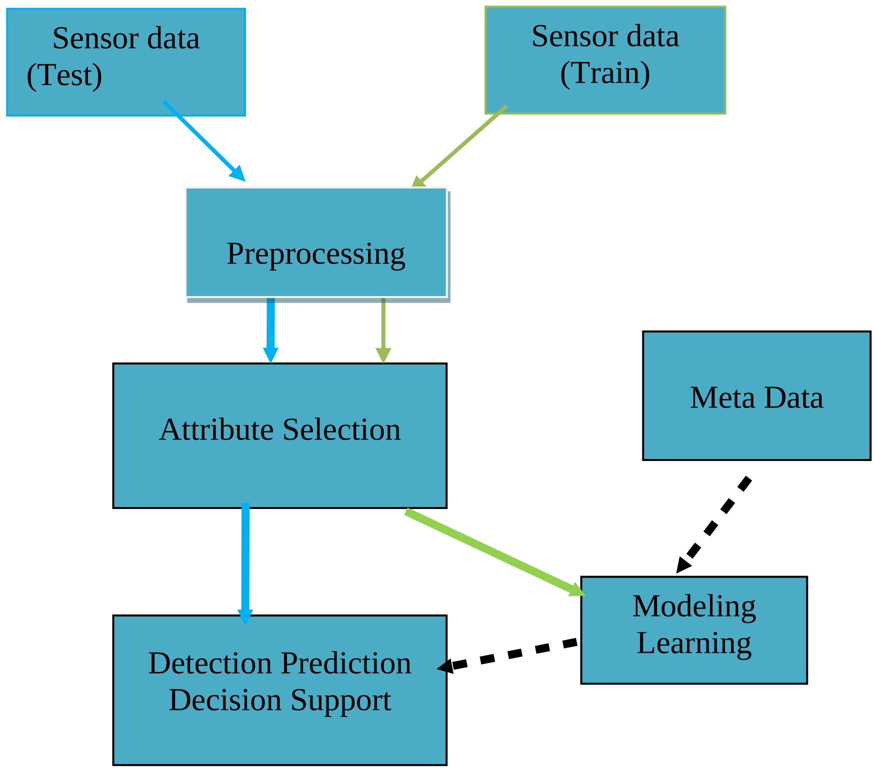


Figure 2.2: The main DM approach for wearable sensor data

It should be noted that other parameters of data mining and machine learning methods are important such as expert knowledge, historical data measurements, electronic health records, and stable parameters (e.g., sex, age). This metadata provides contextual analysis and improves the process of knowledge extraction . The details of the main steps in the data mining approach are presented below:

- **Data Set**

The type of input data and its properties is the prerequisite of any data processing system in order to handle the significant issues such as: selecting the proper data mining approach, designing and adjusting new method.

Several input sources and data gathering methods have been considered in the literature for wearable sensor data in health monitoring systems. Here, three major data gathering approaches have been identified such as experimental wearable sensor data, clinical or online databases of sensor data, and simulated sensor data.

i. Experimental wearable sensor data

Researchers have developed health monitoring systems mostly using their own data gathering experiments to design, model and test the data analysis step . In this case the gathered data are usually obtained based on the pre-defined scenarios due to the testing and evaluation of the performed results .

ii. Clinical or online databases of sensor data

Several studies in this area have used the stored clinical data sets. Developed data mining methods is defined and designed for wearable health monitoring systems, but to evaluate quantitatively and test the performance of output decision of the framework, the most of the works used categorized and complex multivariate data sets with formal definitions and annotations by domain expert . Very common example of online databases is the database. Several papers in the literature have used two main data sets in the PhysioBank.

iii. Simulated sensor data

For the sake of having a wide controlled analysis system, few works have designed and tested their data mining methods through shapely simulated wearable sensors data . Data simulation would be useful when focus of data processing method is on the efficiency and robustness of information extraction . Another reason to create and use simulated data is the lack of long term and large-scale data sets , which helps the proposed data mining systems to deal with huge amount of data.

- **Preprocessing**

Preprocessing of raw data in healthcare is necessary due to the occurrence of noise, motion artifacts, and sensor errors in any wearable sensor networks. Preprocessing in the healthcare domain involves (1) filter unusual data to remove artifacts and (2) remove high frequency noise . The main challenges of the preprocessing phase in healthcare systems are addressed in which includes data formatting, data normalization, and data synchronization. However, there is no tailored work considering these issues in detail for real life scenarios. Since the gathered sensor data is often unreliable and massive.

- **Attribute Selection (AS)**

Generally, for mining massive and real world data sets, the abstraction of raw data in any data mining approach is a way to design and build a model in order to retrieve valuable information. The aim of **AS** is to discover the main characteristics of a data set, which are identical representatives of the original data . Especially in wearable sensor data, according to the magnitude and complexity of the raw data, feature extraction provides a meaningful representation of the sensor data, which can formulate the relation of raw data with the expected knowledge for decision making . Moreover, reducing the amount of sensor network data is another task in feature extraction and **AS** phases, which leads to have an arranged vector of Attribute as an input of **DM** techniques like classifier methods .

- **Modeling and Learning Methods**

There are several **DM** methods and machine learning techniques used in analyzing sensors data in **AmI** such like Neural networks, fuzzy rules, Decision making, and spatial-temporal reasoning Support Vector Machines**,** Decision Trees, Rule-Based Methods, Statistical Tools and ensembles models and others methods. These methods and techniques can help accomplish many important tasks in **AmI** assisted healthcare monitoring and make the system more efficient .

- **Data Mining Tasks**

In the healthcare wearable sensors monitoring there are three main **DM** tasks: prediction, anomaly detection which may include the subtask of raising alarms, and diagnosis where a decision making process is made to often categorize the data into different groups depending on the diseases.

2.4 Information Architecture(IA)

The term information architecture (IA) is coined by Richard Saul Wurman to describe the need to transform data into meaningful information for people to use. . **IA** is later defined by Brancheau and Wetherbe as a high level map of information requirements of an organization or a process of architecting information in order to achieve organizational benefit. . There are many definitions, views and conceptualizations of **IA** depending on its application. In , **IA** is defined as a professional practice and field of studies focused on solving the basic problems of accessing and using, the big amounts of information available today. In , **IA** represents a higher level of abstraction, emphasizing an awareness of systems in terms of how critical subcomponents interacts according to semantic aspects of processes, designs, and metrics. In , **IA** have been defined as supporting business processes by using methods, techniques and software to design, control and analyze operational processes involving humans, organizations, applications, documents and other sources of information . **IA** represents a higher level of abstraction, emphasizing an awareness of systems in terms of how critical subcomponents interact according to semantic aspects of processes, designs, and metrics . In , **IA** is used to organize information about a topic in order to manage it in a structured way.

2.4.1 Enterprise architecture framework (EAF)

The term Enterprise Architecture Framework (EAF) is mostly used to specify a list of important abstraction mechanisms such as perspectives, viewpoints, and dimensions. Thus an **EAF** is a documentation structure for Enterprise Architectures. In , **EAF** is defined as a kind of implicit conceptual Meta model of the architecture of their information technology (IT) systems. It describes the architecture of a business and its **IT**, and their alignment. The term **EAF** is mostly used to specify a list of important abstraction mechanisms such as perspectives, viewpoints, and dimensions. Thus an **EAF** is a documentation structure for Enterprise Architectures. There are a number of reference architectural frameworks including; The Open Group Architectural Framework (TOGAF), Department of Defense Architecture Framework

(DoDAF), Federal Enterprise Architecture Framework (FEAF), Treasury Enterprise Architecture Framework (TEAF) and Zachman Framework.

The **TEAF** creating a matrix of four view (columns) and four perspectives (rows) compared to the one suggested by Zachman . **TOGAF** is a high level and holistic approach to the design of enterprise architecture. It covers four architectural domains; Business, Application, Data, and Technology, which can be seen as views . **TEAF** was developed to promote interoperability. Its dimensions are entities (what), activities (how) and locations (where). With respect to the views it provides five level perspectives for more details . **DoDAF** builds on three sets of views; operational, system and technical standards Operational view describes the operational elements, and tasks. Systems view describes systems and interconnections to support the Operational View. Technical Standards describes rules governing the arrangement, interdependence of system components to augment the Systems View .

In the study conducted by Goethals, . The purpose of the study was to present a direct comparison of the frameworks, based on his views and aspects. According to the authors they studied several existing enterprise architecture frameworks, which helped them to establish a common ground for the framework comparison. Comparison was made based on the perspectives of their stakeholders and abstractions. Finally the research concluded, that Zachman framework appears to be the most comprehensive using a number of viewpoints (dimensions and perspectives) relate to different aspects. "*Zachman framework is a generic reference model, which serves as the basis for numerous other models e.g.* ***TOGAF****,* ***FEAF*** *and* ***TEAF***" .

2.4.2 Zachman Framework of Architecture

Zachman, who is recognized as an expert on Enterprise Architecture (EA), introduced a well-defined framework of architecture having strong and logical connection between business processes, organization strategies and **EA**. This is considered to be one of the major origins of the field of **EA** . Zachman describes the aim of this framework as an architecture that represents the information systems' artifacts, providing a means of ensuring that standards for creating the information environment exist and they are approximately integrated . This framework was first introduced by Zachman in 1987 and was called Information Systems Architecture Framework which then was extended in . Originally the Information Systems Architecture Framework proposed by Zachman had only three aspects Data, Function, and Network. In the extended framework which was then named the Enterprise Architecture Framework by Zachman and Sowa, three more columns or aspects of the enterprise were added namely People, Time, and Motivation which represented the business aspects of the enterprise. According to the authors the **ZF** can also be defined as a conceptual methodology, which shows how all of the specific architectures that an organization might define can be integrated into a comprehensive and coherent environment for enterprise systems. It is an analytical model that organizes various representations of architecture. The **ZF** is a two-dimensional classification schema, a normalized schema. It is the intersection between two historical classifications that have been in use for literally thousands of years, the universal linguistic communications classification of primitive interrogatives: What, How, Where, Who, When, and Why; and the classification of audience perspectives: Owner, Designer, Builder, bounded by the Scoping perspective, and the Implementation perspective **.**

Table 2.3. Zachman framework

	Content (What)	Function (How)	Network (Where)	People (Who)	Time (When)	Motivation (Why)
Scope (Contextual) Planner	List of things important to business.	List of core business processes	List of business locations	List of important organizations	List of Events	List of business goals/ Strategies
Business model (Conceptual) Owner)	Conceptual data / object model	Business Process Model	Business Logistics System	Work Flow Model	Master Schedule	Business Plan
System model (Logical) Designer	Logical Data Model	System Architectu re Model	Distributed systems architecture	Human Interface architecture	Processing Structure	Business Role Model
Technology model(Physical) Builder	Physical Data /Class Model	Technolog y Design Model	Technology Architecture	Presentation Architecture	Control Structure	Rule Design
Detailed Representations (out-fcontext) Subcontractor	Data Definitions	Program	Network Architecture	Security Architecture	Timing Definition	Rule Specificatio n
Functioning Enterprise	Usable Data	Working Function	Usable Network	Functioning Organization	Implemented Schedule	Working Strategy

ZF is typically depicted as a 6 x 6 matrix in which the architecture is described using two independent aspects, rows represent the different audience perspective used to view a business, and the columns represent the various communication interrogatives which apply to each perspective of the business

The **ZF** of **EA** is one of the most widely accepted frameworks amongst the other **EAF**. As Zachman proposed the purpose of this framework is to provide a logical structure which classifies and organizes the descriptive representations of an enterprise that are significant to the management of the enterprise as well as to the development of the enterprise's systems. Zachman has provided different Perspectives for different rows of his proposed framework.

The following Perspectives are depicted by different rows in **ZF**.

- Scope (Planner's Perspective): the planner is concerned with defining the context for the enterprise including specifying its scope.
- Business Model (Owner's Perspective): the owner is interested in modeling the enterprise using business modeling techniques yielding business deliverables.
- System model (Designer's Perspective): the designer had to ensure that the enterprise is so modeled that it fulfills the owner's expectations. He tries to logically model the IT environment.
- Technology Model (Builder's Perspective): the builder is responsible for assembling and managing the various components of the system. The logical design models developed by the designer are mapped onto technology dependent design models to give rise to physical models.
- Detailed Representations (Subcontractor's Perspective): the subcontractor has to manufacture out-of-context components for meeting the builder's expectations. He is responsible for the detailed implementation models.
- Functioning Enterprise: this includes the real working enterprise. Columns of the **ZF** provide focus on each of the perspective while keeping others constant. They facilitate the abstraction of the enterprise's information in a way that is suitable for modeling purposes .

- Data (What?): this column answers the question, ‘What are the important things that the enterprise is dealing with?’ It gives the material composition of the object, the bill-of-materials for enterprises, the data models .
- Function (How): the question, ‘How does it run?’ is answered by the function column. The rows in this column describe the translation process of the mission of an enterprise into more detailed objectives.
- Network (Where): this aspect is concerned with the geographic locations where the enterprise’s activities are distributed.
- People (Who) : it tries to answer the question, ‘Who does what work?’ So this aspect describes who all are involved in the business and what are their functions.
- Time (When) : this aspect tries to answer the question, ‘When do things happen relative to one another?’ It describes the effects of time on the enterprise’s business.
- Motivation (Why) : the question, ‘Why the enterprise does what it does?’ is answered by this aspect. This domain is concerned with the translation of the enterprise’s strategies into specific objectives.

According to Varga, , the purpose of **ZF** is to provide a basic structure that supports the organization, access, integration, interpretation, development, management, and changing of a set of architectural representations. There are certain rules that govern the framework, which provides the framework’s integrity. These rules of the framework are summarized as following:-

Rule 1: Do not add rows or columns to the Framework
Rule 2: Each column has a simple generic model
Rule 3: Each cell model specializes its column’s generic model
Rule 3: Corollary a: Level of detail is a function of a cell, not a column
Rule 4: No meta concept can be classified Into more than one cell
Rule 5: Do not create diagonal relationships between cells
Rule 6: Do not Change the names of the rows or columns
Rule 7: The logic is generic, recursive

As we have seen earlier there are various frameworks for enterprise architecture developed after the **ZF** for enterprise architecture. According to Pereira and Sousa, , the **ZF** is the most widely known framework in the Enterprise Architecture context. It is the most referenced framework, which makes itself a basis for evaluating, establishing, and customizing other enterprise architectural frameworks, methods, and tools. . The reason for its extensive popularity and use is that it is an extremely flexible framework and just defines the logical structure of any enterprise.

Thus it does not impose a particular method or any restrictions on users to use a particular set of pre-defined artifacts unlike other frameworks developed in this field. noted that various organizations, do their enterprise architecture related activities upon the **ZF**, which is by far the highest rate amongst all the other frameworks. Although the US FEAF is gaining popularity amongst these organizations, but FEAF has been developed using the **ZF** as a basis and influence.

2.4.3. Strengths of Zachman Framework

The Zachman Framework of architecture is the most popular framework in the domain of Enterprise Architecture. It is also considered a basis for many other frameworks developed after the **ZF** such as FEAF. According to Zachman, this framework for **EA,** which was formerly known as the framework for information systems architecture, has proven quite valuable for .

- Placing a wide variety of tools and methodologies in relation to one another.
- Improving the communications within the information systems community.
- Understanding the reasons for developing any architectural representation.
- Understanding the risks of not developing any architectural representation.
- Rethinking the classic approach of “application development process”.

Also as pointed out by Fatolahi and Shams, , most of the **EA** tools such as System Architect have compatibility with **ZF**. Along with this the most applied and used methodology for **EA** planning is also intended to develop its products based on **ZF** for architecture.

2.4.4. Weaknesses of Zachman Framework

Although the **ZF** is amongst the most popular frameworks of architectures in the field of **EA** but it has some drawbacks which researchers have shown concern for in the past. The **ZF** is very generic and can over simplify some of the enterprise issues such as its business performance and behavior, although it takes into consideration decision support systems, analytical processing and data exploration. . Some researchers have argued in the past that it is not an easy task to build up architectures using the **ZF** for architecture. Since the framework is firmly constrained using rigorous formal rules, which govern the framework's integrity some difficulties appear in building up architectures if a full coverage on the framework is intended. , have summarized these difficulties in three major problems:

- A lack of methodology covering all the aspects of the framework.
- A lack of repository storing the framework in accordance with the integrity rules.
- Lack of a popular modeling notation for all of the framework's columns.

2.5. Summary of the Chapter

This Chapter presented research works and viewpoints concerning the main theme of the research. Accordingly, literature on the concept of **AmIHCM**, and **DM**/**ML** approaches were discussed. It is learned from the literature that traditional **HCM** is an escalating problem for medical domain and the society that needs consistent research efforts.

According to the literature review there are huge requirements to move the routine medical check and healthcare services from the traditional healthcare monitoring to a new paradigm. Thus **AmI** for healthcare monitoring and personalized healthcare is a promising solution to provide efficient medical services, which could significantly lower down the healthcare cost and improve the **HCM**. In addition, it is mentioned that there are different efforts from different perspectives so far. **DM**/**ML** approach was also discussed to get an understanding of its application, which exhibits its versatility in medical domain.

Also according to the literature review there are a number of reference architectural frameworks that can be used to construct **AmI** healthcare monitoring architecture framework.

The **ZF** based IA is the most popular framework in the domain of **EA**. It is also considered as a basis for many other frameworks developed after the **ZF**. It is evident that **ZF** can be used to define architecture for any object or system of interest. Accordingly, this thesis presents **ZF** to study and describe **AmIHCM** and its management.

3 RELATED WORKS

3.1 Overview

This Chapter is dedicated to the discussion of healthcare monitoring problems in data collection, analysis and dissemination. The chapter focuses on healthcare monitoring Architecture/ Applications and **DM/ ML** techniques used in **AmI** healthcare monitoring. The process of reviewing related works revealed some interesting facts, which provided useful insight in the consideration of the proposed architecture for **AmI** healthcare monitoring information management.

3.2. AmI healthcare Architectures and Applications

AmIHCM systems are much more complicated than traditional computing systems. The development of **AmI-** based software requires creating increasingly complex and flexible applications, so there is a trend toward reusing resources and share compatible platforms or architectures .

As **AmI** started to capture attention during the last ten years, there are several frameworks and Architectures proposed by many researchers. Until recently, the majority of projects had their focus set on achieving functionalities for very concrete environments such as architectures , taking ad-hoc design and implementation decisions focused on the particular environment dramatically reducing the possibility of adapting those projects to other environments, objectives or technologies. However, to operate in a real life scenario, every one of those projects needed solutions for very similar problems.

Another architecture approach, which has been widely used by the researchers is the multi-agent paradigm (MAS). It is based on the division of the system into multiple autonomous components, called agents, that collaborate to resolve functionality. The key characteristic is agent autonomy, as it implies that the agents must present other characteristics that make them very attractive to **AmI**. THOMAS architecture , basically consists of a set of modular services.

THOMAS feeds initially on the FIPA architecture , and it expands its capabilities to deal with organizations, and to boost its services abilities. THOMAS is specifically addressed to design organizational structures for multi-agent systems. Agents have access to the THOMAS infrastructure through a range of services included on different modules or components. Agents that are capable of autonomous decision making, incorporate learning mechanisms, and are able to respond to events by planning and preplanning in execution time. THOMAS is an open architecture that can easily incorporate any type of agent.

presented MaRV architecture that has evolved from the THOMAS architecture by Bajo and Corchado, to facilitate the integration of agents and smart wearable devices via wireless networks and mobile technology. The MaRV MAS is based on a belief, desire, and intention (BDI) model have been developed by Corchado et al., , in which the agent's function as controllers and coordinators for various medical care tasks. The MaRV MAS is a specialized feature of the THOMAS architecture for intelligent environments that can address the need to improve techniques for obtaining resident and patient data, as well as assign diagnoses in hospital centers and geriatric facilities, and monitor all types of patients. The agents can initiate services on demand, or according to planned actions. The MaRV MAS is a distributed agent platform that uses a WSN to establish remote communication between patients and caregivers. All smart wearable devices in MaRV are based on RFID technology. Recent years have given way to a number of multi-agent architectures that utilize data merging to improve their output and efficiency.

presented Alzheimer multi-agent system (ALZ-MAS), which is a MAS aimed at enhancing the assistance and healthcare for Alzheimer patients. The main functionalities in the system include reasoning and planning mechanisms. That is embedded into the agents, and the use of several context aware technologies to acquire information from users and their environment. However these systems must provide common improvements such as knowledge discovery, and knowledge representations.

developed Telemonitoring homecare aimed at enhancing remote healthcare of dependent people at their homes. The main contribution of this development is the use of an experimental architecture developed by Tapia et al., , that allows the interconnection of heterogeneous WSNs (i.e. multiple technologies) and is based on the **AmI** paradigm. This architecture formalizes the integration of services, communications and wireless technologies to automatically obtain information from users and the environment in an evenly distributed way, focusing on the characteristics of ubiquity, awareness, intelligence and mobility.

presented CodeBlue developed at Harvard University. CodeBlue is hardware and software platform. The hardware design part includes the design and development of a mote-base pulse oximeter, two-lead ECG, and a motion analysis sensor board. The software architecture is based on a publish/subscribe routing framework. Code Blue aims to provide coordination and communication among wireless medical devices in an ad hoc manner. The sensors do not publish data at an arbitrary rate, because the wireless channel's bandwidth is limited and they filter the data locally. Moreover, when publishers and subscribers are not within radio range, multi-hop routing is used. Since the publishers and subscribers are mobile.

Also, a discovery protocol is used for Code Blue nodes to discover each other and determine the capabilities of their sensor devices. Moreover, the system integrates a localization system called MoteTrack. However these model are for specific purpose and must provide common improvements such as service discovery, rich knowledge representations and context-awareness.

In recent years, a lot of researchers have been working on the definition and development of architectures that deal with the complex characteristics of **AmI** environments. However, only small parts of the literature review try to formally define conceptual models that are not restricted to particular AmI domains or environments.

Some recent research efforts presented by Rui et al., defend the need of general reference models in AmI and present a conceptual hierarchical model to describe a typical **AmI** application system. They proposed **AmI** systems using a five-layer model: sensors and actuators layer, AmI network and middleware layer, devices layer, service layer and **AmI** applications layer. However, they do not include, as part of this work, high-level development tools that could closely guide developers in the process of building systems that fulfill the proposed conceptual model. Other research is the PERSONA project , which focuses on the specification of a reference architecture for ambient assisted living (AAL) spaces. The authors formally define a logical architecture that abstracts all functionality not leading to context, input, or output events as services.

Others researchers presented their projects with a general purpose in mind. These projects provided high-level abstractions and communications capabilities suitable for the typical requirements of **AmI** systems. Examples are the AMIGO project , focused on the home entertainment environment. , provided an implementation of a general purpose AmI framework.

Others researchers used simulated sensor data in their projects.. As stated by , the reason to create and use simulated data is the lack of long term and large scale data sets, which helps the proposed data mining systems to deal with huge amount of data. Also according to Wang et al., Zhu, . Data simulation would be useful when more focus of data processing method is on the efficiency and robustness of information extraction, rather than handling real-world data.

simulated the environment of Baraha Medical City in Shambat, Khartoum North, in Sudan using the framework reported in . The Hospital receives patients who suffer from chronic diseases such as heart diseases, asthma, diabetes and abnormal blood pressure etc. Also people in post-surgery state needs continuous monitoring of their health condition, especially the vital signs, until their health status becomes stable. In the work by Salih and Abraham , the focus was only on the monitoring patients vital signs and providing medical service for patients with chronic or terminally ill diseases. Depending on the critical condition of the patent, each patient was attached with several sensors. In the project, the main task is to develop Novel Intelligent Ensemble Health Care Decision Support and Monitoring System that could assist the hospital management to assess the situation of the hospital as Normal or Abnormal (too many medical emergencies) so that more medical help could be sorted.

AmI has potential applications in many areas of life, including in the home, healthcare system, elderly, transport, and industry, safety systems, and supported living of many different variations. In this research we focus on healthcare and elderly systems. HEARTFAID project , which aims to defining efficient and effective health care. developed ALARM-NET, which is an

Assisted-Living and Residential Monitoring Network for pervasive, adaptive healthcare developed at the University of Virginia. It integrates environmental and physiological sensors in a scalable, heterogeneous architecture, which supports real-time collection and processing of sensor data. Communication is secured end-to-end to protect sensitive medical and operational information.

In the SAPHIRE project developed by Laleci et al., , the patient monitoring is achieved by using agent technology complemented with intelligent decision support systems based on clinical practice guidelines . The patient monitoring is achieved by using agent technology complemented with intelligent decision support systems based on clinical practice guidelines. The observations received from wireless medical sensors together with the patient medical history will be used in the reasoning process. The patient's history stored in medical information systems will be accessed through semantically enriched Web services to tackle the interoperability problem. In order to guarantee long term patient monitoring and successful execution of clinical decision support systems.

3.3. Data Mining and Machine Learning Techniques Used in AmI

The past few years have an increase in the development of wearable sensors for health monitoring systems. This increase has been due to several factors such as development in sensor technology as well as the society, which address the need for providing new methods for care given increasing challenges with an aging population. An important aspect of investigation in such system is how the data is treated and analyzed.

As discussed in Chapter Two, a number of data mining-related studies have been undertaken to analyze **AmI** healthcare monitoring data, with results frequently varying depending on the conditions and the technique used. As indicated in Akyildiz et al., , when analyzing sensor data, **AmI** systems may employ a centralized or distributed model. Sensors in the centralized model transmit data to a central server, which fuses and analyzes the data it receives. In the distributed model, each sensor has onboard processing capabilities and performs local computation before communicating partial results to other nodes in the sensor network. The choice of model will have a dramatic effect on the computational architecture and type of sensor that is used for the task as described by Benini and Poncino, and also by Jayasimha et al., . In both cases, sensor data is collected from disparate sources and later analyzed to produce information that is more accurate, more complete, or more insightful than the individual pieces.

There are several **AI** and **DM** methods and techniques used in analyzing sensors data in **AmI** such like Neural networks, fuzzy Rules, Reasoning, Decision making, and spatial-temporal reasoning and machine learning. These methods and techniques can help accomplish many important tasks in **AmI** assisted **HCM** and make the system more efficient .

Artificial Neural Network (ANN) is widely used for classification and prediction. proposed an **ANN** based activity recognition system in order to determine the occurrence of falls. Their system works with single sensor placed on to the chest of the subjects. However **ANN** required more tuning parameters than support vector machines, and also **ANN** is sensitive to noise (a validation set may help here) and missing values in the training samples need to be replaced or removed. Also presented a multi-layered feed forward neural network (FNNs) as activity classifiers and recognized 8 daily activities with an overall performance of 95%.

used fuzzy logic-based techniques to learn user preferences in an ordinary living environment. They have devised an experimental intelligent inhabited environment, called the iDorm (intelligent dormitory) at the University of Essex, UK. The iDorm contains space for various activities such as sleeping, working, and entertaining, and contains various items of furniture such as bed, desk, etc. The sensors can sense temperature, occupancy. iDorm deals with two types of rules, static (user-independent) rules such as how to react in an emergency, and to lower the lights and temperature when the Rooms is available , and learned rules reflecting the user preferences. The learning uses a fuzzy logic-based technique called Incremental Synchronous Learning (ISL). After monitoring and learning, the iDorm agent can take control of the environment. If the user behavior changes, the learning system may need to modify some of the rules, so it will go back to learning phase in which there can be a repeated learning process.

Augusto and Nugent used Event-Condition-Action (ECA) Rules and various extensions of them for applications in Smart Homes and supported living for the elderly. The intuitive reading of such rules is that on detecting certain events, if certain conditions are true then certain actions should be executed. The event part (first line) is the trigger of the **ECA** rule. The rule is triggered if an event occurs that matches the event part of the rule. Then if the condition (second line) of a triggered rule is true the rule fires, requiring execution of the action (third line). An example of such is the work has been done by Augusto et al., is by looking into the use of **ECA**. These are proposed to fulfill two criteria. Monitoring of patients for safety, long term monitoring of patients for profiling and learning behavioral patterns.

developed another Smart Home application and area of testing, for the Medical Automation Research Center. This system uses probabilistic methods to determine patterns in behavior. Based on a series of sensors, one in each room, the system monitors the duration of time that the user spends in each room. Although these systems have shown improvements over other systems of its type, it is still lacking in one major area. Both the systems deal with the elderly living alone and there is no identifier on the person using the system,

developed single person home for the elderly, which assumes no partner, no visitors, no health care providers and no maintenance people entering the house. Any one else entering the house will cause the systems to gather false data.

developed GerAmI system in conjunction with the Alzheimer Santísima Trinidad Residence of Salamanca, an institute with multiple stories, multiple rooms and upwards of 40 residents. As with all previously mentioned for **AmI** systems, the GerAmI uses sensors to record patient and user data. However rather than sensors using motion or heat to track users, each resident and staff wears a bracelet containing a unique **RFID**. As each bracelet's **RFID** is unique it allows all of the residents and staff to be tracked individually without false data being recorded.

This system is unique in that it also tracks the movements of the staff members. This is a major benefit in a system such as this when the medical care providers are on hand as it allows faster reactions to emergencies by alerting staff that are on duty and also located closer to the source of the problem. If intervention or assistance is required a message is sent to the staff members personal digital assistant (PDA). The message contains the name or identifier of the patient in question, the problem that has occurred as well as information from the database about the best way to deal with the situation based on previous events.

Other’s research studies related to Sensing and acting provided links between **DM** algorithms and environment. A number of methods of reasoning must take place to make such algorithms responsive, adaptive, and beneficial to users. These include user modeling, activity prediction and recognition, decision-making, and spatial- temporal reasoning. The most common data source for model building is low-level sensor information. This data is easy to collect and process. However, one challenge in using such low-level data is the voluminous nature of the data collection.

developed the MavHome smart home project and collected motion and lighting information in an average of 10,310 events each day. In their project, a **DM** pre-processor identifies common sequential patterns, then uses the patterns to build a hierarchical model of resident behavior. relied upon this sensor data to determine the resident action and device state, and then pulls information from similar situations to provide a context-aware environment.

developed the iDorm research and focuses on automating a living environment. However, instead of a Markov model, they model resident behavior by learning fuzzy rules that map sensor state to actuator readings representing resident actions. The amount of data created by sensors can create a computational challenge for modeling algorithms. However, the challenge is even greater for researchers who incorporate audio and visual data into the resident model. used video data and inter-transaction (sequential) association rules in resident actions.

Activity Prediction and Recognition is widely used in **DM** and **ML** field that helps to identify events, which have not yet occurred. Some researchers have used offline data analysis in Activity Prediction and Recognition.

used accelerometer data to identify different kinds of activities like walking, running, climbing stairs. All these approaches rely on offline data analysis to learn “typical” contexts. proposed a method based on Kohonen Self Organizing Maps (KSOMs) and k-Means clustering, which is able to identify typical motion profiles. This approach relies on active training, used to construct a supervised context transition profile based on a first order Markov process to make the **KSOM** training procedure converge, the neighborhood radius of the learning neurons must decrease over time . However **KSOM** have strong dependence of the initialization and is has too unbalanced classes, and also K- Means clustering has problems when clusters are of differing sizes, densities, non-globular shapes, and empty clusters.

presented critical events that can be detected using classification algorithms, for which Bayes classifiers are known to provide good results. However, traditional classifiers do not allow meaningful interruption until the entire model has been evaluated, which is crucial in mobile devices due to limited resources. Limited processing power and high data rates limit the time available for processing one set of sensor values. To overcome this limitation they employed a novel anytime Bayes classifier in a two-phase architecture. On the back-end server a full index structure is stored, which is an extension of previous work was presented by Assent et al., for anytime stream classification. It is trained by sequences of sensor measurements, which correspond to normal situations.

implemented Support Vector Machines (SVM's) to predict clinical scores of the severity of data obtained from wearable sensors in patients with Parkinson's disease. SVM's have the ability to generate nonlinear decision boundaries, by mapping the feature space into a higher dimensional space (using kernels) where classes are linearly separable.

presented an **AmI** application, which is focused on a single environment, which has outfitted with sensors and designed to improve the experience of the resident in the environment., presented the Neural Network House, and the MavHome was presented by . developed projects to adaptively control home environments by anticipating the location, routes and activities of the residents (i.e., a person moving within an **AmI** space).

developed prediction algorithms as multiple types. developed predicting resident locations, and even resident actions that allow the **AmI** system to anticipate the resident's needs and assist with (or possibly automate) performing the action. The modeling techniques described so far can be characterized as unsupervised learning approaches. However, if resident activity data is available, then supervised learning approaches can be used to build a model of resident activity and use it to recognize observed activities.

employed a naive Bayes learner to identify resident activity from among a set of 35 possible classes, based on collected sensor data. However Naïve Bayes is simple probabilistic classifier based on the assumption that the features for a given class are mutually independent, which means that the decisions are made as if all features are equally important. enhanced the model with object interaction data.

Over the last few years, supporting technologies for **AmI** have emerged. Automated decision-making and control techniques are available for Building a fully automated **AmI** application. discussed how **AI** planning systems could be employed to not only remind individuals of their typical next daily activity, but also to complete a task if needed. described the use of temporal reasoning with a rule-based system to identify hazardous situations and return an environment to a safe state while contacting the resident. Few fully implemented applications decision-making technologies have been implemented. also have used a reinforcement learner to automate physical environments with volunteer residents, in the MavPad apartment and the MavLab workplace. presented the iDorm, which is another of these notable projects that uses fuzzy rules learned through observation of resident behavior. These rules can be added, modified, and deleted as necessary, which allows the environment to adapt to changing behavior. However, unlike the reinforcement learner approaches, automation is based on imitating resident behavior and therefore is more difficult to employ for alternative goals such as energy efficiency.

employed a Hierarchical Task Network (HTN) planner to generate sequences of actions and contingency plans that will achieve the goal of the **AmI** algorithm. For example, the **AmI** system may respond to a sensed health need by calling a medical specialist and sending health vitals using any available device (cell phone, email, or fax). If there is no response from the specialist, the **AmI** system would phone the nearest hospital and request ambulance assistance. , has developed novel computer systems enhancing the quality of life of people suffering from Alzheimer's disease and similar disorders, that help an individual perform daily tasks by sensing the individual's location and environment, learning to recognize patterns of behavior, offering audible and physical help, and decision making to alerting caregivers in case of danger.

described dynamic sequential decision making in medicine using Markov-based approach originally described in terms of medical decision-making. utilized dynamic influence diagrams. While the authors in [127,128] have utilized decision trees to model temporal decisions.

In all cases, the goal is to determine optimal sequences of decisions. Markov decision processes (MDPs) is an efficient technique for determining optimal sequential decisions (termed a “policy”) in dynamic and uncertain environments have been used by Schaefer et al., and Alagoz et al., .

investigated the experimental results of the performance of different classification techniques for classifying the data from different wearable sensors used for monitoring different diseases. The Base Classifiers Proposed used in this work are K- nearest neighbor(IBK) , Attribute Selected Classifier , Partial Decision Trees(PART) , Decision Tree Algorithm J48 , Logistic Model Trees (LMT) , Random Forest and the Random Tree algorithm . Experiments are conducted on wearable sensors vital signs data set, which was simulated using a hospital environment. The main focus was to reduce the dimensionality of the attributes and perform different comparative analysis and evaluation using various evaluation methods like Error Metrics, ROC curves, Confusion Matrix, Sensitivity and Specificity. Experimental results reveal that the proposed framework is very efficient and can achieve high accuracy.

Several studies have focused on the importance of the ensemble methods in the field of medical health care monitoring. These studies have applied different approaches to the given problem and achieved high classification accuracies. Ensemble methods combined a set of individual methods to obtain a better more accurate and reliable estimates or decisions than can be obtained from using a single model. Classification of sensory data is a major research problem in **WSNs**. Many researchers have utilized ensemble models in **AmI** assisted **HCM**.

presented Classifier Ensemble Optimization method for activity recognition by optimizing the output of multiple classifiers with evolutionary algorithm. They have combined the measurement level output of different classifiers in terms of weights for each activity class to make up the ensemble. Classifier ensemble learner generates activity rules by optimizing the prediction accuracy of weighted feature vectors to obtain significant improvement over raw classification. presented a comparison of single supervised machine learning and ensemble methods in classifying seven publicly available cancerous data. The authors used C4.5 decision tree, bagged decision tree on seven publicly available cancerous micro array data, and compared the prediction performance of these methods. The experimental results indicate that the ensemble methods consistently perform well over all the datasets in terms of their specificity.

A combinational feature selection and ensemble neural network method is introduced by Liu et al., for classification of biomedical data. Many individual algorithms such as self-organizing maps (SOM), learning vector quantization (LVQ), multi-layer perceptron's (MLPs), neural-fuzzy systems, and SVMs were applied to ECG signals. However, these methods have been typically applied to distinguish normal signals from abnormal signals across patients. This is difficult because of the substantial variation in the morphologies of ECG signals across patients. For this reason, Li et al., implemented an ensemble consisting of a standard SVM designed to distinguish normal signals from abnormal signals across patients and a set of one-class SVMs, presented by Scholkopf et al., , (one per patient) to distinguish normal signals for a given patient from all other signals .

proposed the use of bagging with C4.5 algorithm, bagging with Naïve bayes algorithm to diagnose the heart disease of a patient. Other study used bagging algorithm to identify the warning signs of heart disease in patients and compared the results of decision tree induction with and without bagging. used Naive Bayes, J48 Decision Tree and Bagging algorithm to predict the survivability for Heart Diseases patients.

conducted experiments on ECG data to identify abnormal high frequency electrocardiograph using decision tree algorithm C4.5 with bagging. Ensemble methods have been used by others researchers in other type of medical data sets. presented a new classification algorithm TBWC combination of decision tree with bagging and clustering. experimented on ovarian tumor data to diagnose cancer-using C4.5 with and without bagging. experimented on breast cancer data using C4.5 algorithm with bagging to predict breast cancer survivability.

Table 3.1: Evaluation of DM methods conducted in AmIHCM .

Authors	Data Mining Methods	Solution Objective	Weakness of the methods

	Neural Network	Activity recognition system	ANN Require more tuning parameters than support vector machines.
	FNNs	Activity classifiers and recognize	Multilayer FNNs need enough training samples and hidden nodes to be able to approximate any function.
	Fuzzy logic-based	Learn user preferences in an ordinary living environment	Difficult to employ for alternative goals such as energy efficiency.
, ,	ECA	Smart Homes and supported living for the elderly.	Systems deal with the elderly living alone. There is no identifier on the person using the system.
	KSOMs) and k-Means clustering	Activity Prediction and Recognition	KSOM have strong dependence of the initialization and unbalanced classes, and also K- Means clustering has problems when clusters are of differing sizes, densities, non-globular shapes, and empty clusters.
	Bayes classifiers	Critical events that can be detected using classification algorithms	Traditional classifiers do not allow meaningful interruption until the entire model has been evaluated, which is crucial in mobile devices due to limited resources. Limited processing power and high data rates limit the time available for processing one set of sensor values.
	SVM’s	To predict clinical scores of the severity of data obtained from wearable sensors in patients with Parkinson’s disease.	SVM’s take long training times and the decision boundaries can be highly complex and Extension to classification of more than two classes is usually time -consuming.
	Naive Bayes	Identify resident activity from among a set of 35 possible classes	It is assumed that the data attributes are conditionally independent which is not always so. Lack of available probability data is a significant disadvantage of the Naïve Bayesian approach
,	Rule-based	To identify hazardous situations and return an environment to a safe state while contacting the resident.	More difficult to employ for alternative goals such as energy efficiency.
,	Decision trees	Temporal decisions	It is restricted to one output attribute.
	Base Classifiers IBk,Attribute Selected Classifier,	Classifying the data from different wearable sensors used for monitoring different diseases	Ensemble methods combined a set of individual methods to obtain a better more accurate and reliable estimates or decisions than can be obtained from using a single model.

	Bagging, PART, J48, LMT, Random Forest, Random Committee and the Random Tree algorithm		
	C4.5 with bagging. Ensemble methods	Identify abnormal high frequency of ECG electrocardiograph	Does not work well with small training data set. Small variation in data can lead to different decision

developed a Novel Intelligent Ensemble Health Care Decision Support and Monitoring System that could assist the hospital management to assess the situation of the hospital as Normal or Abnormal (too many medical emergencies) so that more medical help could be sorted. Experiments are conducted on wearable sensors vital signs data set, which was simulated using a hospital environment. First, the authors carried out a thorough investigation comparing the performance of various base classifiers. Second, the authors carried out a thorough investigation comparing the performance of various Meta base classifiers. These Meta classifiers used are LogitBoost , Bagging , AdaBoostM1, Random Committee , Stacking , and Voting . Third, we investigated Meta classifiers and new Novel Intelligent Ensemble method was constructed based of Meta classifier Voting combining with three base classifiers J48, Random Forest and Random Tree algorithms. The results obtained show that the Novel Intelligent Ensemble method classifier achieved better outcomes that are significantly better compared with the outcomes of the all Base Classifiers Proposed and all meta base classifiers used in this Thesis. Different comparative analysis and evaluation were done using various evaluation methods like Error Metrics, ROC curves, Confusion Matrix, Sensitivity, Specificity and the Cost/Benefit methods. The results obtained show that the Novel Intelligent Ensemble method classifier is very efficient and can achieve high accuracy and, better outcomes that are significantly better compared with the outcomes of the all base classifiers proposed and all meta base classifiers.

Others researchers used spatial and temporal reasoning to have better understanding of the activities in an **AmI** application. Spatial and temporal reasoning are two well-established areas of AI as stated by Galton, . They have been the subjects of intense research for a couple of decades and there are well-known formalisms and algorithms to deal with spatial, temporal, and spatial-temporal reasoning. has shown how the traditional frameworks for spatial reasoning and for temporal reasoning can be used to have a better understanding of the activities in an **AmI** application. used such a language to integrate both concepts in the same formalism and to obtain spatial temporal reasoning combined with active databases in the identification of interesting situations. presented an alternative formalism for reasoning about time based on Allen's temporal logic.

Summary of the Chapter

The Chapter focuses on reviewing the architectures and data mining methods used in the literature for applications involving **AmI HCM** systems. The survey of literature exhibited that systematic and enterprise view is lacking in **AmI** healthcare monitoring information management. Though, there is good understanding of the importance of **AmI** healthcare information systems by various authors, their focus was limited to a single aspect of the whole information system. Systemic and integrated architecture view is lacking towards addressing the problem and it still is short of an architectural framework that guides the development and management of **AmI** healthcare monitoring data collection and analysis systems. Also, in the literature, there is still a lack of a commonly accepted architecture to build the **AmI** systems of the future, with predictable dependability properties. This issue is considered in , where the author pointed out the concepts of “architecture" and “system" need to be redefined in the context of **AmI** systems, in order to properly define “ambient dependability" attributes, threats and means.

Also the chapter focus is put on the data mining and machine learning algorithms that have been used in order to provide an overview of the algorithm's capabilities and shortcomings. Many such researchers focus on giving a global overview of the topic most of them are related to general studies for healthcare i.e., well known problems in healthcare with simple and routine data mining approaches . Table 3.1 summarized the key findings and evaluation of **DM** methods Conducted in **AmIHCM**.

DM techniques that consider the specific challenges, which emerges from data coming from wearable sensors is of ever increasing importance. In **HCM** systems focus has been recently shifting from that of obtaining data to one of developing intelligent algorithms to perform a variety of the tasks . Such tasks not only include traditional pattern recognition and anomaly detection but also must consider decision-based systems, which can handle context awareness, and subject specific models and personalization. These latter challenges are particularly important when healthcare monitoring services are designed to address the need of decision support **HCM** system, such as distributed health monitoring and long-term prevention.

DM techniques that have been applied to **HCM** data have varied and it is also not uncommon that several techniques are used within the same architecture. These systems must provide common improvements such as service discovery, rich knowledge representations and context-awareness. Analysis are limited at least in a local context and no noble model ensemble and trend analysis was conducted so far to explore the variance in the determinant factors and improvements in accuracy of the analysis models in **AmI HCM** system. It is also noted that **DM** using classification is accepted and is an appropriate approach in the **HCM** domain.

4 RESEARCH METHODOLOGY

4.1 Overview

This Chapter presents and discusses the approach and methods of the research. Hence, it covers the methodological aspects that have guided the present work. The Chapter starts with research approach and Theoretical Frameworks. The qualitative approaches of this research are described next, followed by, the machine learning techniques which are mainly used to determine the information requirement of the AmIHCM information management process. The Chapter concludes by the evaluation of the research process.

4.2 Research Approach and Theoretical Frameworks

There are two research paradigms that characterize much of the research in the IS discipline, as mentioned by : behavioral science and design science. The behavioral science paradigm seeks to develop and verify theories that explain or predict human or organizational behavior. The design-science paradigm seeks to extend the boundaries of human and organizational capabilities by creating new and innovative artifacts. In the design-science paradigm, knowledge and understanding of a problem domain and its solution are achieved in the building and application of the designed artifact.

The design-science paradigm has its roots in engineering and sciences . It is fundamentally a problem-solving paradigm. It seeks to create innovations that define the ideas, practices, technical capabilities, and products through which the analysis design, implementation, management, and use of information systems can be effectively and efficiently accomplished . According to , research methods can be classified in mainly two ways, quantitative research methods and qualitative research methods.

Quantitative methods were originally developed in the natural sciences to study natural phenomena. These methods include surveys, laboratory experiments, formal methods like econometrics, and numerical methods like mathematical modeling. Qualitative research methods were first developed in the subjects of social sciences to help the researchers to study the social and cultural phenomena. These include action research, case study method, ethnography, grounded theory, and phenomenology. The author chose to use a case study research method in this Thesis. Case study research is the most common qualitative method used in information systems . It is employed in this research as it enables to bring the researchers to an understanding of a complex issue or object and can extend experience or add strength to what is already known through previous research. It is used to identify the information requirement of the **HCM** domain and define enterprise wide information architecture.

This research is descriptive in method, because its main purpose is to explain and define **HCMIA** through conceptual modeling by utilizing **HCM** data collection and analysis practices. On the other hand, it is fundamental in purpose because it **HCMIA** based on theoretically genuine, contextual and empirical models.

This research combines literature review, **DM**, **ML** and a qualitative research approach, in a design science research paradigm, in order to explain healthcare monitoring situations in Khartoum state hospitals (Sudan) and defines information architecture for **AmIHCM** information management. Literature on **AmIHCM**, **DM**, **ML** and **IA** provides an excellent insight to the concept, evolution need, guiding frameworks use and application **of IA**, **DM** and **ML** techniques are used to investigate the simulation wearable sensors patients monitoring data situations through identification of patients as normal or abnormal. A qualitative research approach is used to define **AmIHCM** information management requirements. Regarding the general process; with background Literature Review – Related Work, motivation, problem statement and objectives given, Figure 4.1 depicts major phases of this research. As shown in the Figure, the research work commenced with the background Literature review. The next task was qualitative HCM data collection and analysis, which was done in parallel with the analysis of simulation vital signs patient's monitoring situations through machine learning techniques including trend analysis to develop novel ensemble model. In the next stage, the researcher uses the ensemble model obtained with the HCM data collection to define and developing the integrated **AmIHCMIA**. Finally the research findings and **AmIHCMIA** were evaluated based on appropriate validation techniques. Conclusions, recommendations and indication to future research are the last important tasks of the research process.

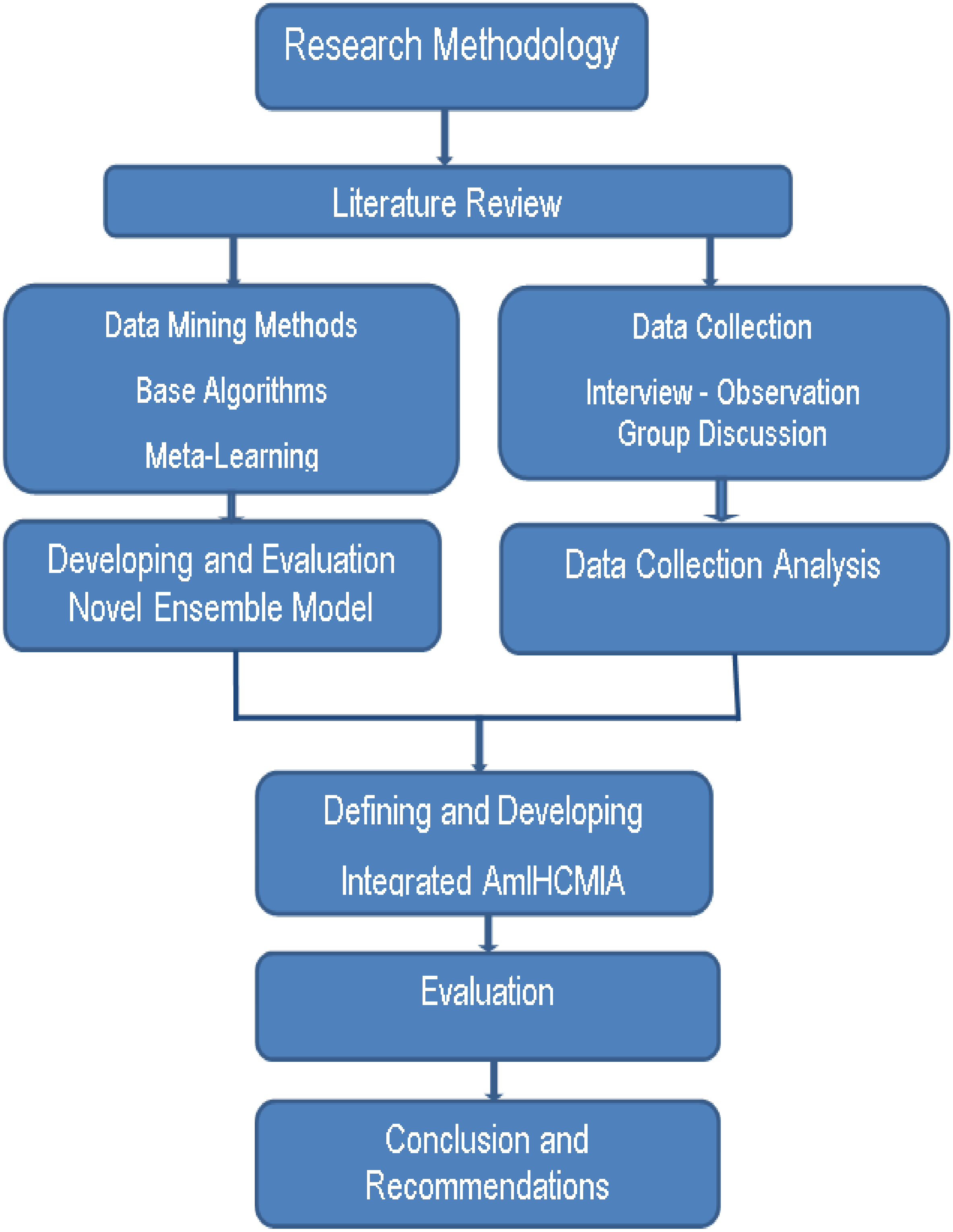


Figure 4.1: Conceptual model of the research process

4.3 Qualitative Data Collection and Analysis Approaches

As indicated previously, the research is based on seven private hospitals and eighteen public hospitals were selected randomly in Khartoum state (Sudan). The detail of the process is presented in this Section.

4.3.1 Study Population and Sampling

The unit of analysis is the **HCM** of inquiry data system with patients monitoring data reporting, analysis and dissemination practice and activities being the phenomenon. Accordingly, the population research has been based on **HCM** in Twenty five selected hospitals in Khartoum state (Sudan). Seven private and eighteen public hospitals were selected randomly, also at the ministry of health in Khartoum State. This makes the research more comprehensive and representative; the majority of the registered patients are concentrated at the capital Khartoum State.

The sampling frame comprises all stakeholder organizations working in the departments of **HCM** data in hospitals. This also includes experts in healthcare monitoring, head of information system in these hospitals and in the Ministry of health in Khartoum state. With recognizable variations, literature in the qualitative research area suggests sample sizes ranging from 3 to 30 . Though there appear no agreement on sample sizes, based on recommendations and discussions in qualitative research literature, 17 respondents were selected from organizations in the area. The purposive convenience sampling was adopted, which is the most frequently used approach in qualitative studies **.**

4.3.2 Data Collection Techniques

In the effort of addressing the research problem, various empirical data regarding different **HCM** data aspects needs to be collected. In doing so, there are different qualitative techniques of data collection. Qualitative data sources include observation and participant observation, interviews and questionnaires, documents and texts, and the researcher's impressions and reactions . Written data sources can include published and unpublished documents like reports and newspaper articles. In qualitative study, use of different data collection technique is highly recommended in order to get a full picture of the subject at hand, which is **HCM** in the context of this research.

Accordingly, the preferred data collection techniques for this specific research are interviews, observation, focused group discussion, and document analysis. Particularly, interview and focused group discussion were used to collect the required data from healthcare monitoring experts in selected hospitals, while observation and document analysis helps for closer understanding of the case.

4.3.3 Data Analysis procedures

According to literature in qualitative research, it is not easy to clearly draw a line between data collection and analysis tasks. The analysis affects the data and the data affect the analysis in significant ways. Therefore, literature prefers to speak of "modes of analysis" rather than "data analysis" in qualitative research. These modes of analysis are approaches to gathering, analyzing and interpreting qualitative data. The common thread is that all qualitative modes of analysis are concerned primarily with textual analysis, whether verbal or written . The analysis of this research is based on the principle of the hermeneutic circle, which suggests that a deeper understanding of a text or text-analogue **HCM** information management) in relation to its context can only take place through a back and forth movement of renewed understandings .

In doing so, the interview sessions, repeated observation and focused group discussions helped a lot. As stated by Darke et al., , the three concurrent activities in data analysis, namely data reduction, data display and conclusion drawing /verification were employed. Data reduction refers to the process of selecting, simplifying, abstracting and transforming raw case data. Data display refers to the organized assembly of information using techniques like narratives, tables, matrices etc, to enable drawing of conclusions. Conclusion drawing /verification involves drawing meaning from data and building logical chain of evidence. Cross-analysis of the healthcare monitoring data collection and analysis practices in the organizations was mapped against the Zachman Framework to determine the dimensions and elements of the architecture. With its variety of modeling profiles.

The Unified Modeling Language (UML) is chosen as a modeling tool to document architectural artifacts.The **UML** is a standard language for writing software blueprints. The **UML** may be used to visualize, specify, construct, and document the artifacts of a software intensive system. The **UML** is appropriate for modeling systems from enterprise information systems. To apply the **UML** effectively starts with forming a conceptual model, which requires three major elements: the **UML** basic building blocks, the rules that dictate how these building blocks may be put together, and common mechanisms that apply throughout the language . A diagram is the graphical presentation of a set of elements, most often issued as a connected graph of vertices (things) and arcs (relationships). Drawing diagrams to visualize a system from different perspectives, so a diagram is a projection into a system. There are several types of **UML** diagrams such as, Object diagram, Use case diagram, Sequence diagram, Collaboration diagram, Statechart diagram, Activity diagram, Component diagram, Deployment diagram.**UML** has been used effectively for such domains such as, enterprise information systems, banking and financial services, telecommunications, transportation, medical electronics .

The **UML** and its business profile could be freely used for **EA** modeling based on the Zachman framework as discussed in and . Literature providing conceptual frameworks and review of internal practices are also part of the effort and approaches in the process of the study.

4.4 Machine Learning / Data Mining Techniques

As investigating healthcare monitoring situations through wearable sensors data analysis is one of the major concerns of this research, data mining techniques were used in the process of identifying a novel approach in discovering potential knowledge hidden in wearable sensors patients data accumulated. Data mining is the analysis of observational data sets to find unsuspected relationships and summarize the data in novel ways that are both understandable and useful to the data owner . Researchers tried to prove its applicability in many domain areas and organizations. One of such areas could be healthcare monitoring system, where very big patients monitoring data is accumulated. However this data is in long hand written format in majority of hospitals and the in others hospitals it is still in a process of transferring healthcare monitoring patients data to a computer system. Thus for lack of long term data sets which helps the proposed data mining systems the simulation is used to deal with huge amount of data. In addition high costs and time is needed to build big dataset.

According to Wang et al., Zhu , data simulation would be useful when the focus of data processing method is on the efficiency and robustness of information extraction. Another reason to create and use simulated data mentioned by Zhu , is the lack of long term and large scale data sets, which helps the proposed data mining systems to deal with huge amount of data.

In this research, the data set was constructed by simulation of patients wearable sensors from the environment of Baraha Medical City in Shambat, Khartoum North, Sudan. Using the GerAmi framework reported in , a platform is built in Java and C# with GUI compatibly with mobile phones and personal computers.   The machine learning methodology used was guided by .

Then, follows exploration of data quality issues, pre-processing and attribute selection tasks relevant to the data mining goal, model building and evaluation along with a possible recommendation. With respect to the specific techniques employed, classification methods were used in this Thesis. First, various single classification algorithms were selected thorough experimental investigations by comparing the performance of various base classifiers to construct the base classifiers.

Second, the author carried out a thorough investigation comparing the performance of various Meta base classifiers. Third, the author investigated Meta classifiers and a new Novel Intelligent Ensemble method was constructed based of Meta classifier Voting combining with three base classifiers.

A brief description of the single classification algorithms, Meta base classifiers and the model combination techniques used in this Thesis is presented in the next subsections.

4.4.1 Decision Trees

A “divide-and-conquer” approach to the problem of learning from a set of independent instances leads naturally to a style of representation called a decision tree . Nodes in a decision tree involve testing a particular attribute. Usually, the test at a node compares an attribute value with a constant. However, some trees compare two attributes with each other, or use some function of one or more attributes. Leaf nodes give a classification that applies to all instances that reach the leaf or a set of classifications, or a probability distribution over all possible classifications.

To classify an unknown instance, it is routed down the tree according to the values of the attributes tested in successive nodes, and when a leaf is reached the instance is classified according to the class assigned to the leaf. If the attribute is numeric, the test at a node usually determines whether its value is greater or less than a predetermined constant, giving a two-way split. Alternatively, a three-way split may be used, in which case there are several different possibilities.

The divide-and-conquer was developed and refined over many years by J. Ross Quinlan of the University of Sydney, Australia . Although others have worked on similar methods, Quinlan’s research has always been at the very forefront of decision tree induction. The method that has been described using the information gain criterion is essentially the same as one known as ID3 (Iterative Dichotomiser 3).

The use of the gain ratio was one of many improvements that were made to ID3 over several years. described it as robust under a wide variety of circumstances. Although a robust and practical solution, it sacrifices some of the elegance and clean theoretical motivation of the information gain criterion. A series of improvements to ID3 culminated in a practical and influential system for decision tree induction called C4.5. These improvements include methods for dealing with numeric attributes, missing values, noisy data, and generating rules from trees.

4.4.2 Decision Tree Algorithm J48

J48 classifier is a simple C4.5 decision tree for classification. developed C4.5 algorithm, which is used to generate a Decision Tree. It creates a binary tree. The decision tree approach is most useful in classification problem. With this technique, a tree is constructed to model the classification process.

Once the tree is built, it is applied to each tuple in the database and results in classification for that tuple . The C4.5 unlike the IDE3, accepts both continuous and categorical attributes in building the decision tree. It has an enhanced method of tree pruning that reduces misclassification errors, due to noise or too-much detail in the training data set. Decision Trees are produced from the J48 i.e. Open Source Java implementation of C4.5 release in . This is a standard Decision Tree algorithm.

The basic Algorithm is presented as follows in .

- Check for base cases
- For each attribute "a" find the normalized information gain from splitting on "a"
- Let a_best be the attribute with the highest normalized information gain
- Create a decision node that splits on a best
- Recourse on the sub lists obtained by splitting on a best, and add those nodes as children of node

4.4.3 Partial Decision Trees (PART)

Two well-known members of the family of rule-learners are C4.5 and RIPPER . Both approaches perform two steps to induce their rule sets. First, an initial rule set is determined, and second, these rules are adjusted or discarded according to a global optimization strategy. C4.5, for instance, generates an un-pruned decision tree and transforms this tree into a set of rules. For each path from the root node to a leaf, a rule is generated. Then, each rule is simplified separately followed by a rule-ranking strategy. Finally, the algorithm deletes rules from the rule set as long as the rule set's error rate on the training instances decreases. RIPPER implements a divide- conquer strategy to rule induction. Only one rule is generated at a time and the instances from a training set covered by this rule are removed. It iteratively derives new rules for the remaining instances of the training set. Frank and Witten describe a rule induction approach without the need for applying a global optimization strategy to generate appropriate rules . **PART** (Partial Decision Trees) adopts the divide-and-conquer strategy of RIPPER and combines it with the decision tree approach of C4.5 .

More precisely, PART generates a set of rules according to the divide-and-conquer strategy, removes all instances from the training collection that are covered by this rule and proceeds recursively until no instance remains. To generate a single rule, **PART** builds a partial decision tree for the current set of instances and chooses the leaf with the largest coverage as the new rule. Afterwards, the partial decision tree is discarded, which avoids early generalization.

4.4.4 Logistic Model Trees (LMT)

Logistic Model Trees consist of a decision tree structure with logistic regression function at the leaves . As in decision tree, the tested attributes is associated with every inner node. The attributes with *k* values, the node has *k* child nodes for nominal attributes and depending on the value of the attribute the instances are sorted down. For the numeric attributes, the node has two child nodes and comparing the attributes of tested value to a threshold (the instances are sorted down based on threshold . **LMT** uses pruning of cost complexity. Compared to other algorithm, it is slower to compute.

4.4.5 Logit Boost algorithm

The LogitBoost algorithm was introduced by Friedman et al., . The algorithm is similar to AdaBoost, with the main difference being that LogitBoost performs stage wise minimization of the negative binomial log likelihood, while AdaBoost performs stage wise minimization of the exponential loss. By virtue of using the binomial log likelihood instead of the exponential loss, the LogitBoost algorithm was believed to be more “gentle” and consequently likely to perform better than AdaBoost for classification problems in which the Bayes error is substantially larger than zero.

4.4.6 Random Forest

Random Forest developed by Breiman, is a group of un-pruned classification or regression trees made from the random selection of samples of the training data. Random features are selected in the induction process. Prediction is made by aggregating (majority vote for classification or averaging for regression) the predictions of the ensemble. Each tree is grown as described in **:** By Sampling N randomly, if the number of cases in the training set is N but with replacement, from the original data. This sample will be used as the training set for growing the tree. For M number of input variables, the variable m is selected such that m<< M is specified at each node, m variables are selected at random out of the M and the best split on these m is used for splitting the node. During the forest growing, the value of m is held constant. Each tree is grown to the largest possible extent. No pruning is used. Random Forest generally exhibits a significant performance improvement as compared to single tree classifier such as C4.5. The generalization error rate that it yields compares favorably to Adaboost, however it is more robust to noise.

Random forest is an ensemble classifier that consists of many decision trees. The output of the classes is represented by individual trees. The random forest inducing algorithm is derived from random decision a forest that was proposed by Tin Kam Ho of Bell Labs in 1995 . This method combines random selection of features to construct a decision tree with controlled variations. The tree is constructed using algorithm as detailed below.

i) Let N be the number of training classes and M be the number of variables in classifier.
ii) The input variable m is used to determine the node of the tree. Note that m<M.
iii) Choosing n times of training sets with the replacement of all available training cases N by predicting the classes, estimate the error of the tree.
iv) Choose m variable randomly for each node of the tree and calculate the best split.
v) At last the tree is fully grown and it is not pruned.

The tree is pushed down for predicting a new sample. When the terminal node ends up the label is assigned the training sample . This procedure is iterated over all trees and it is reported as random forest prediction.

4.4.7 Random Tree

A random tree is a tree constructed randomly from a set of possible trees having *K* random features at each node. "At random" in this context means that in the set of trees each tree has an equal chance of being sampled. Or we can say that trees have a "uniform" distribution. Random trees can be generated efficiently and the combination of large sets of random trees generally leads to accurate models. There has been an extensive research in the recent years over Random trees in the field of machine Learning.

A random tree is a tree formed by stochastic process. Types of random trees include Uniform spanning tree, Random minimal spanning tree, Random binary tree, Random recursive tree, Treap, Rapidly exploring random tree, Brownian tree, Random forest and branching process .

4.4.8 K- nearest neighbor IBK

IBK (commonly known as K- nearest neighbor). instance-based learning is lazy , deferring the real work as long as possible, whereas other methods are eager, producing a generalization as soon as the data has been seen. In instance-based learning, each new instance is compared with existing ones using a distance metric, and the closest existing instance is used to assign the class to the new one.

This is called the nearest-neighbor classification method. Sometimes more than one nearest neighbor is used, and the majority class of the closest k neighbors (or the distance-weighted average, if the class is numeric) is assigned to the new instance.

This is termed the k-nearest-neighbor method. Computing the distance between two examples is trivial when examples have just one numeric attribute: it is just the difference between the two attribute values. It is almost as straightforward when there are several numeric attributes: generally, the standard Euclidean distance is used.

4.5 Meta-learning

The idea of Meta-learning is to execute a number of concept learning processes on a number of data subsets, and combine their collective results through an extra level of learning. Meta-learning aims to compute a number of independent classifiers by applying learning programs to a collection of independent and inherently distributed databases in parallel. The “base classifiers” computed are then collected and combined by another learning process. The most popular meta-learning algorithms are bagging and boosting. Bagging , is a method for generating multiple classifiers (learners) from the same training set. The final class is chosen by, e.g., voting.

Combining multiple models approach is to making decisions more reliable is to combine the output of different models. Several machine learning techniques do this by learning an ensemble of models and using them in combination: prominent among these are schemes called bagging, boosting, and stacking. They can all, more often than not, increase predictive performance over a single model. And they are general techniques that can be applied to numeric prediction problems and to classification tasks.

4.5.1 AdaBoostM1

AdaBoost.M1is a well-known algorithm for boosting weak classifiers . AdaBoostM1 is a member of a broader family of iterative machine learning algorithms that build the final classifier through a finite series of improvements to the classifier. AdaBoost.M1 is the most straightforward generalization of boosting algorithm. It is adequate when the weak learner is strong enough to achieve high accuracy.

4.5.2 LogitBoost

One of the boosting algorithms developed recently, is introduced for predicting protein structural classes. Logit Boost is one of the boosting algorithms developed in recent years. Boosting was originally proposed to combine several weak classifiers to improve the classification performance. Later on, Freund and Schapir proposed a more capable and practical boosting algorithm, the so-called AdaBoost. . Ada- Boost, an abbreviation for Adaptive Boosting, is a metal earning algorithm. It tries to build a weak classifier iteratively on others according to the performance of the previous weak classifiers.

4.5.3 Bagging

The term bagging (for “bootstrap aggregating”) was coined by , who investigated the properties of bagging theoretically and empirically for both classification and numeric prediction. . Combining the decisions of different models means amalgamating the various outputs into a single prediction. The simplest way to do this in the case of classification is to take a vote (perhaps a weighted vote); in the case of numeric prediction, it is to calculate the average (perhaps a weighted average). Bagging and boosting both adopt this approach, but they derive the individual models in different ways. In bagging, the models receive equal weight, whereas in boosting, weighting is used to give more influence to the more successful ones—just as an executive might place different values on the advice of different experts depending on how experienced they are.

Bagging (bootstrap aggregating), generates a collection of new sets by resampling the given training set at random and with replacement. These sets are called bootstrap samples. New classifiers are then trained, one for each of these new training sets. They are amalgamated via a majority vote. .

4.5.4 Stacking

Stacked generalization , originated with , who presented the idea in the neural network literature, and was applied to numeric prediction by . Compared different level-1 models empirically and found that a simple linear model performs best. Authors also demonstrated the advantage of using probabilities as level-1 data. A combination of stacking and bagging has also been investigated .

Stacked generalization, is a different way of combining multiple models. Although developed some years ago, it is less widely used than bagging and boosting, partly because it is difficult to analyze theoretically and partly because there is no generally accepted best way of doing it—the basic idea can be applied in many different variations.

Unlike bagging and boosting, stacking is not normally used to combine models of the same type. The usual procedure would be to estimate the expected error of each algorithm by cross validation and to choose the best one to form a model for prediction on future data. But isn’t there a better way? With three learning algorithms available, can’t we use all three for prediction and combine the outputs together? One way to combine outputs is by voting- the same mechanism used in bagging.

4.5.5 Random Committee

Classifier that ensembles randomizable base classifiers, it builds an ensemble of base classifiers and averages their predictions. Each one is based on the same data but uses a different random number seed. This only makes sense if the base classifier is randomized; otherwise, all classifiers would be the same. The random committee algorithm is a diverse ensemble of random tree classifiers. In the case of classification, the random committee algorithm generates predictions by averaging probability estimates over these classification trees .

4.6 Ensemble methodology

The main purpose of an ensemble methodology is to combine a set of models, each of which solves the same original problem, in order to obtain a better composite global model with more accurate and reliable estimates or decisions than can be obtained from using a single model. The main discovery is that the ensemble classifier is constructed by ensemble machine learning algorithms, such as bagging and boosting approaches, often performs much better than the single classifiers that make them up.

The idea of ensemble methodology is to build a predictive model by integrating multiple models. It is well known that ensemble methods can be used for improving prediction performance. The learning procedure for ensemble algorithms can be divided into the following two parts. .

1. Constructing base classifiers/base models: the main tasks of this division are:
   (a) Data processing: Prepare the input training data for building base classifiers and attributes selection to reduce the dimensionality of the attributes.
   (b) Base classifier constructions: build base classifiers on the data set with a learning algorithm.
2. Voting: the second stage of ensemble methods is to combine the base classifiers models built in the previous step into the final ensemble model.

4.6.1 Voting

There are various kinds of voting systems. Two main voting systems are generally utilized, namely weighted voting and un-weighted voting. In the weighted voting system, each base classifier holds different voting power. On the other hand, in the unweight system, individual base classifier has equal weight, and the winner is the one with most number of votes. The simplest kind of ensemble is the way of aggregating a collection of prediction values base level giving different voting power for its prediction. The final prediction obtains the highest number of votes. Voting includes the weighted average (of each base classifier holds) when using regression problem and majority voting when doing classification and the weighted-majority output is given by:

$$\arg max \left(\sum_{i=1}^{k} P_i\,(x), w_i \right) \qquad (1)$$

Pi(x) is the results of the prediction of i th prediction model and Pi(x, w) is indicator function defined as:

$$P_i(x, w) = \begin{cases} 1 & x = w \\ 0 & otherwise \end{cases} \qquad (2)$$

Majority voting has some benefits that it does not require any additional complex computation and any previous knowledge. However, this approach leads to the result that it is difficult to analyze and interpret. The second strategy is un-weighted, which gives some predictor higher weight if they achieve more accuracy than others (the winner is the one with the most number of votes) .

Combining rules are the simplest combination approach and it is probably the most commonly used in the multiple classifier system . This combination approach is called non-trainable combiner, because combiners are ready to operate as soon as the classifiers are trained and they do not require any further training of the ensemble as a whole .

A theoretical framework for fixed rules combination was proposed by . It includes the sum, product, max, min, average and median rules. In this Thesis we have used the Maximum rule. Maximum rule is based on information provided by the maximum value of :

$$P\left(x^i \mid w_k\right) \qquad (3)$$

Across all class labels, it finds the maximum score of each class between the classifiers and assigns the input pattern to class with the maximum score among the maximum scores as following .

$$f(x) = w_j, j = \arg\max\left(\max p\left(x^i \middle| w_k\right)\right) \qquad (4)$$

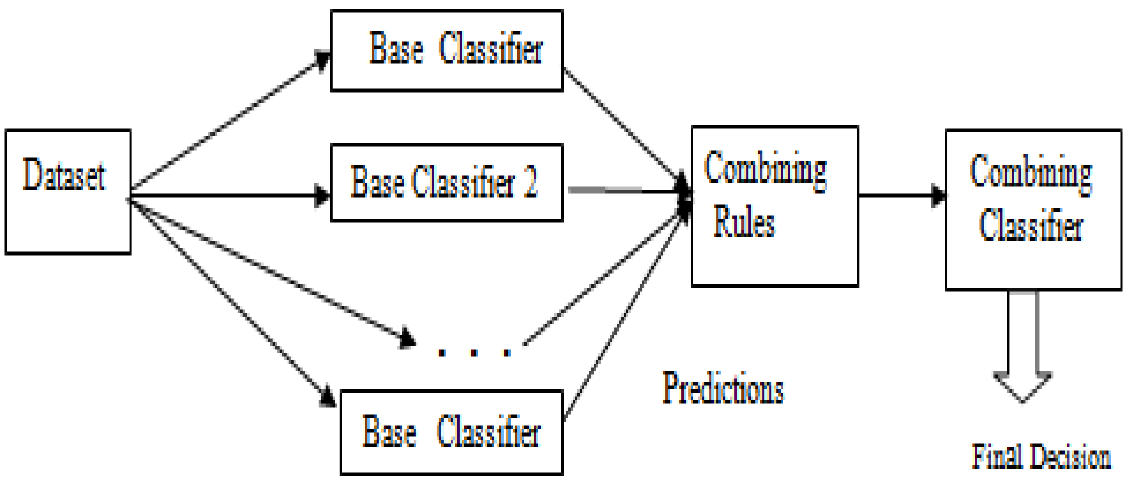


Figure 4.2: Ensemble using combination rule with voting

As shown in Figure 4.2, the dataset (which are simulation sensors data in our case) are used to train and test the system, each classifier in the system is trained using the training data set, and then give an output. The outputs of all classifiers are combined using one of fixed rules that mentioned previously to give the final decision.

In this Thesis the author investigated Meta classifiers and a new Novel Intelligent Ensemble method was constructed based of Meta classifier Voting combining with three base classifiers J48 , Random Forest and Random Tree algorithms.

4.7 Cross-validation

The holdout method reserves a certain amount for testing and uses the remainder for training (and sets part of that aside for validation, if required). Instead, you should ensure that the random sampling is done in such a way as to guarantee that each class is properly represented in both training and test sets. This procedure is called stratification, and we might speak of stratified holdout . A simple variant forms the basis of an important statistical technique called cross-validation. In cross-validation, you decide on a fixed number of folds, or partitions of the data.

In this Thesis the author applied a 10-fold cross validation test option. Cross-Validation (CV) is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments: one used to learn or train a model and the other used to validate the model. The basic form of **CV** is k-fold **CV**. In k-fold **CV** the data is first partitioned into k equally (or nearly equally) sized segments or folds. Subsequently k iterations of training and validation are performed such that, within each iteration a different fold of the data is held-out for validation while the remaining k -1 folds are used for learning. The advantage of K-Fold Cross validation is that all the examples in the dataset are eventually used for both training and testing .

4.8 Attribute selection

Attribute selection (AS), also called feature selection. It is often an essential data processing step prior to applying a learning algorithm. Reduction of the attribute dimensionality leads to a better understandable model and simplifies the usage of different visualization technique. **AS** is the process of identifying and removing as much irrelevant and redundant information as possible. Reduces the dimensionality of the data, may allow learning algorithms to operate faster and more effectively and, accuracy can be improved later on future classification. It finds minimum set of attributes such that resulting probability distribution of data classes is as close as possible of original distribution.

Methods used for **AS** can be classified into two types. The filter approach and Wrapper approach The filter approach actually precedes the actual classification process. The filter approach is independent of the learning algorithm, computationally simple fast and scalable. Using filter method, **AS** is done once and then can be provided as input to different algorithms. .

Wrapper approach uses the method of classification itself to measure the importance of attribute set, hence the **AS** depends on the algorithm model used. Wrapper methods are too expensive for large dimensional database in terms of computational complexity and time since each attribute set considered must be evaluated with the classifier algorithm used.

Filter methods are much faster than wrapper methods and therefore are better suited to high dimensional data sets. Some of these filter methods do not perform attribute selection but only attribute ranking hence they are combined with search method when one needs to find out the appropriate number of attributes. Such filters are often used with forward, backward elimination, bi-directional search, best-first search, and other methods .

Various **AS** techniques have been proposed in the ML/DM literature. In this Thesis, we used **WEKA** tool for pre-processing experiments, to reduce the attributes dimensionality and formulated a new dataset, which was derived from the original dataset after implementing several **AS** algorithms, Such as:

- Correlation-based Feature Selection (CFS)

CFS is a filter algorithm that ranks feature subsets according to a correlation based heuristic evaluation function.**CFS** assumes that useful feature subsets contain features that are predictive of the class but uncorrelated with one another. **CFS** computes a heuristic measure of the “merit” of a feature subset from pair-wise feature correlations and a formula adapted from test theory. Heuristic search is used to traverse the space of feature subsets in reasonable time; the subset with the highest merit found during the search is reported .

- Principal Component Analysis (PCA)

PCA technique reduces the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables .

- Gain Ratio (GR) attribute evaluation

GR is a modification of the information gain that reduces its bias. **GR** takes number and size of branches into account when choosing an attribute. It corrects the information gain by taking the intrinsic information of a split into account. Intrinsic information is entropy of distribution of instances into branches (i.e. how much info do we need to tell which branch an instance belongs to). Value of attribute decreases as intrinsic information gets larger .

- Relief Attribute Evaluation

The main idea of Relief algorithm , is to evaluate and estimate the quality of attributes according to distinguishing values between the instances that are near to each other. Both Relief and its extension ReliefF , are aware of the content information and can correctly estimate the quality of attributes in classification tasks with strong dependencies between attributes .

Some of these filter methods do not perform **AS** but only feature ranking hence they are combined with search method when one needs to find out the appropriate number of attributes. Such filters are often used with forward selection, which considers only additions to the attribute subset, backward elimination, bi-directional search, best-first search, genetic search and other methods. **.**

4.9 Evaluation Approaches and Techniques

Both the process and end result of research undertaking needs to be validated. According to , the utility, quality, and efficiency of a design research need to be rigorously demonstrated via well-executed evaluation methods. The selection of evaluation methods must be matched appropriately with the designed research and the selected evaluation metrics.

Literature shows that several different evaluation techniques exist. As summarized by , these techniques can be categorized broadly as either questionnaires, or measuring techniques (such as simulation or experimentation with a running system).

As this research at hand is a design science research and the subject of this work is a model artifact, appropriate evaluation techniques from the paradigm were used. This agrees with the classification of evaluation methods in design science researches of information systems by as action research, case study, experiments, prototyping and surveying.

Models are evaluated with respect to their fidelity with real world phenomena, level of detail, robustness, and practical utility . Accordingly; surveying, descriptive evaluation, experiment and the case study itself are examined and employed for the evaluation of the design of this research.

Surveying through questionnaire is an evaluation method to confirm the validity, completeness and utility of a model. According to Cleven et al., , by conducting a survey, information is collected through interviewing representatives of a certain target group in the process of evaluating design artifacts.

A descriptive evaluation method involves informed argument using information from relevant researches to build a convincing argument for the research utility. An experimental evaluation method involves the use of the models either in a controlled environment for qualities or in a simulation form to test the usability of it.

In line with the above discussion, questionnaire items reflecting relevant metrics are adapted and modified to collect data to confirm the validity of the architectural model, which is completed by healthcare monitoring experts at hospitals.

It is basically used to conduct interviews with healthcare monitoring experts who are working directly on healthcare monitoring data collection and analysis.

In addition, descriptive evaluation technique, which is an explanation based on literatures is also used as an evaluation method .

Thus, while the models of healthcare monitoring information architecture are evaluated using surveying, descriptive evaluation and the case study itself, simulation wearable sensors data analysis models are evaluated using experimental approach through a training, validate and testing methods in terms of accuracy performance, error rate and the ROC curve.

In this research the author evaluated the performance of the algorithms by measuring their performance by various methods and metric. The following are the methods used in our experimental in this Thesis.

A) Evaluation of time to build a model for each classifier.

B) Error Metrics

The following are the different error metrics used to evaluate each algorithm.

I) Mean Absolute Error (MAE):

Is the average difference between the predicted and actual value in all test case, it is the average prediction error . The formula for calculating MAE is given in equation below:

$$\frac{|a1 - c1| + |a2 - c2| + \cdots + |an - cn|}{n} \quad (5)$$

Assuming that the acual output is a, expected output is c.

II) Root Mean Squared Error (RMSE):

Is frequently used values to predict a model or estimation technique through which the values are observed from the being modeled or estimated . The square root of the mean absolute error given an RMSE as follows:

$$\sqrt{\frac{(a_1 - c_1)^2 + (a_2 - c_2)^2 + \cdots + (a_n - c_n)^2}{n}} \quad (6)$$

While estimating the effectiveness and accuracy of data mining technique it is essential to measure the error rate of each method. In the case of binary classification tasks the error rate takes false positive and false negative, components under consideration. Table 4.1 depicts confusion matrix. The **ROC** analysis, which stands for Receiver Operating Characteristics is applied.

Nowadays the **ROC** curves are used especially in medicine where the distinction between the classification the healthy patient as ill and reverse needs to be done . Additionally the **AUC** (Area Under ROC Curve) is taken under consideration. It measures the model's ability to find the difference between two outcomes .

The **AUC** method is called discrimination. In the perfect case the discrimination is equal 1, however in two outcomes has discrimination rate equal 0.5 because the area under diagonal axis (when no model applied). The **ROC** analysis is directly and naturally related to the cost/benefit analysis of diagnostic decision support . Table 4.1 depicts confusion matrix.

Table 4.1: confusion matrix

Predicted Class	Actual class P	N
P	**True Positive TP**	**False Positive FP**
N	**False Negative FN**	**True Negative TN**
Total	**P**	**N**

I) Accuracy a :

Proportion of the total number of predictions that were correct

$$\frac{(TP + TN)}{(P + N)} \tag{7}$$

II) Precision p :

Proportion of the predicted positive cases that were correct

$$\frac{TP}{(TP + FP)} \tag{8}$$

III) Recall r :

Proportion of positive cases that were correctly identified

$$\frac{TP}{(TP + FN)} \tag{9}$$

IV) F- measures:

Average of the information retrieval precision and recall metrics

$$\frac{p + r}{(2\ pr)} \tag{10}$$

Sensitivity and Specificity

In medical data Sensitivity and Specificity is often used to evaluated and comparing classifiers models.

Sensitivity: proportion of actual positives, which are predicted positive

$$\frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FP})} \tag{11}$$

Specificity: proportion of actual negative, which are predicted negative

$$\frac{\mathrm{TN}}{(\mathrm{TN} + \mathrm{FP})} \tag{12}$$

Another critical issue in validating a research work is a theoretical support both for the process and the output of a research. Accordingly, reference to literature has been used to validate the statements. It is also worth mentioning that during the process of validation of statements, the research has strived to use multiple sources. Member checking through presentations, and repeated observation at the research site, and participatory modes of research were used to ensure reliability and internal validity of the research.

5 DATA ANALYSIS AND EXPERIMENTATION

5.1 Overview

The aim of this Chapter is to investigate simulated wearable sensors monitoring patient's data, and determine the information requirement of **AmIHCM** information management. Accordingly, data mining experiments helps in the reduction of the wearable sensors data set size by removing irrelevant and redundant attributes and in developing novel intelligent ensemble health Care decision support and monitoring system that could assist the hospital management to assess the working situation of the hospital. A qualitative data collection and analysis also helps in determining the information architure requirements of the domain. In line with this, Section 5.2 presents machine learning experimentation and results. The Section 5.3 discusses the result of qualitative data analysis and findings. Section 5.4 presents conclusion formed from the data analysis and experimentation followed by the summary of the Chapter.

5.2 AmIHCM Situation: Experimentation and Results

This Section discusses the findings of machine learning experiments and trend analysis on the simulation wearable sensors patients monitoring data as presented in . The purpose of these experiments are as follows:

Firstly is to investigate the experimental results of the performance of different classification techniques for classifying the simulation wearable sensors patient's data from different wearable sensors used for monitoring different types of diseases. Secondly is to investigate various meta classifiers to develop novel intelligent ensemble combine model for healthcare decision support and monitoring System that could assist the hospital management to assess the situation of the hospital as Normal or Abnormal, so that more medical help could be sorted.

Accordingly, the main research question and two sub questions of this thesis, mentioned in Chapter one and mentioned below, are addressed through the analysis of simulated wearable sensors monitoring data and synthesis of previous experiments.

Does Ambient Intelligence provide efficient medical services, which could significantly improve assisted for healthcare monitoring? If so, in what way (s)?

There are others sub questions:

RQ1: How knowledge discovered from wearable sensor data could be used in practice?

RQ2: How ensemble combine decision model (machine learning) can optimize the results and improve assisted health care monitoring?

In addressing this research question, we define the main objective mentioned in Chapter one as presented below:

- To investigate novel ensemble model by intelligent analysis of simulated wearable sensors patients vital signs monitoring data using data mining tools, to get better results and to improve assisted health care monitoring.

In addition to the main objective, there are sub objectives as following:

- To simulate the patient wearable sensors vital signs monitoring data.
- To investigate the experimental results of the performance of different classification techniques for classifying the data from different wearable sensors vital signs used for monitoring different diseases.
- To evaluate the model with various evaluation methods.

In the process of addressing the objectives of the research indicated above, two major phases of experiments were conducted. In the first phase experiment, attempt has been made to investigate the experimental results of the performance of different classification techniques for classifying the data from different simulated wearable sensors used for monitoring different patients with different diseases. As stated in the methodology section, the Base Classifiers Proposed used in the first experiment are: IBk, Attribute Selected Classifier, Bagging, PART, J48, LMT, Random Forest, Random Committee and Random Tree algorithm. Experiments are conducted on simulation wearable sensors vital signs data set, which was simulated using a hospital environment.

In the second phase experiment, we investigated Meta classifiers and the meta classifiers used in this second experiment are: AdaBoostM1, Bagging, LogitBoost, Random Committee, Stacking, and Voting. Finally new Novel Intelligent Ensemble method was constructed based of Meta classifier voting combining with three base classifiers J48, Random Forest and Random Tree algorithms.

5.3 Data Set and Simulation Environment of Hospital Environment

5.3.1 Software simulation platform

In this research, the researcher used the GerAmi framework reported in , for simulating the Hospital environment of Elbaraha Medical City in Shambat, Khartoum North, Sudan with 30 patients. Platform is built in Java and C# with GUI compatibly with mobile phones and personal computers. Depending on the patient's ailment, different number of sensors was attached and the data is transmitted to a central server using tiny low power sensor devices. The sensor uses a 8 MHz CPU with 10 KB RAM and provided 100 meters range.

Various other sensors are attached depending on the patient's requirement:
Example:
1. Oxymeter: Heart rate and blood oxygen saturation levels with current location
2. ECG: Samples signal 12 bits @ 120 Hz
3. Motion Capture and EMG Sensors: Special-purpose sensors to capture
limb motion and muscle activity.

How the sensors transmit data?
-- Sensors locally filter, compress, or analyze data to reduce radio congestion
-- Data from critical patients given higher priority
Scalable platform to handle 100's of real patients.

5.3.2. Data Set and Simulation of Hospital Environment

We simulated the environment of Baraha Medical City in Shambat, Khartoum North, Sudan using the framework reported in . The hospital is situated in a 600 Square meter lot with a garden within the compound. The hospital has five floors with a 75-bed capacity and provides complete medical services for patients.

The Hospital receives patients who suffer from chronic diseases such as heart diseases, asthma, diabetes and abnormal blood pressure etc. Also people in post-surgery state needs continuous monitoring of their health condition, especially the vital signs, until their health status becomes stable. Each patient when arrived to the hospital with chronic diseases, the patient first received primary healthcare treatment. Then the patient will be assigned to individual room. Figure 5.1 depicts patients with wearable sensors in an hospital environment.

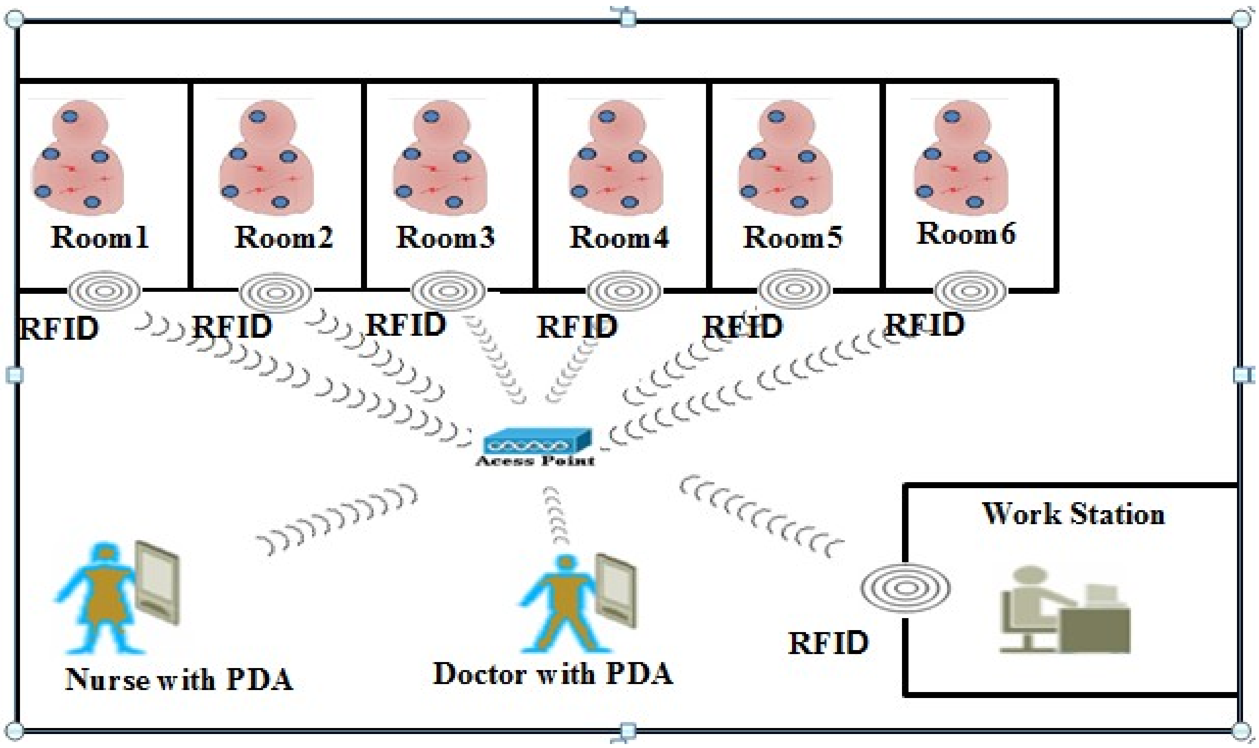


Figure 5.1: Hospital environment

In our simulation, we allocated 6 chronic ill patients in each floor (total 30 patients) as we focused only on the monitoring and providing medical service for patients with chronic or terminally ill diseases. Depending on the critical condition of the patent, each patient was attached with several sensors. For thirty patients, there were a total of 300 readings at any measuring instant.

As the results of the simulation, we obtained the wearable sensors monitoring dataset, A screen shot of sample dataset is presented in (Appendix I).

In this project, our main task is to develop a decision support system that could assist the hospital management to assess the situation of the hospital as Normal or Abnormal (too many medical emergencies) so that more medical help could be sorted. The simulated dataset consist of 745 instances and 300 attributes.

5.3.3 Pre-Processing

Data preparation or preprocessing is always important in machine learning and pattern recognition process. Though, there are various types of preprocessing tasks like handling missing values, minimizing noises, Transformation: changes the forms of the data into the ones appropriate for the data mining task by using different operations, dimensionality reductions etc. Transformation of dataset was done to meet our data-mining task in this research.

In the data set obtained from simulation wearable sensors patient's monitoring data, there is no missing data and also no noises. All the values of the attributes in the Dataset are numeric data and we have changed the last attribute class from numeric to nominal data.

We have class attribute to Abnormal or Normal where a 'Abnormal' specifies 1 class and a 'Normal' Specifies 0 class, also in the process of preprocessing we applied attribute selection so as to reduce the dimensionality of data as presented in the next Subsection.

5.3.4 Attribute Selection (AS)

Attribute Selection (AS) plays an important role in classification. This is one of the Preprocessing techniques in data mining. **AS** is extensively used in the fields of statistics, pattern recognition and medical domain.

AS means reducing the number of attributes. Removing irrelevant and redundant attributes, which do not have significance in classification task, reduces the attributes. The **AS** improves the performance of the classification techniques.

The process of **AS** is as follows:

- Generation of candidate subsets of attributes from original attribute set using searching techniques.
- Evaluation of each candidate subset to determine the relevancy towards the classification task.
- Termination condition to determine the relevant subset or optimal attributes subset.
- Validation to check the selected attributes subset.

We investigated available attribute selection methods (Evaluators) and Search Methods in Weka tool to the Dataset using full training set (300 attributes) . We found that Correlation-based Feature Selection (CFS) Evaluator with Best first, search methods reduce the dimensionality of the attributes to six attributes. Table 5.1summarizes the results.

- Correlation-based Feature Selection (CFS) Evaluator:

Evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Subsets of features that are highly correlated with the class while having low inter-correlation are preferred.

- Best First Search methods:

Searches the space of feature subsets by greedy hill-climbing augmented with a backtracking facility. All 300 attributes were labeled as *A, B, C, Z,.…KN.* We investigated all the available classifiers using WEKA and finally managed to reduce to 6 attributes: AK, CM, CP, CW, FJ and KN.

Table 5.1: Performance of the evaluator and search method used by AS

Evaluator	**Search Method**	**Final No of Attributes**
Correlation-based Feature Selection (CFS)	Best first	6
Correlation-based Feature Selection (CFS)	Greedy Step wise (forwards)	6
Gain Ratio (GR) attribute evaluation	Attribute Ranker	300
Principal Component Analysis (PCA)	Attribute ranking	10
Relief Attribute Evaluation	Attribute ranking	300

5.4 First Phase Experiments and Results

5.4.1 The Base algorithms Proposed

The aim of this Section is to investigate the experimental results of the performance of different classification techniques (see Annex A) for the simulation wearable sensors dataset to select the base classifiers with highest performance accuracy to be used in the second phase in the next section. The performance factors used for analysis are accuracy and error measures. The accuracy measures are TP rate, F Measure, ROC area, Sensitivity and Specificity. The error measures are Mean Absolute Error, Root Mean Squared Error and Kappa Statistics. We investigated algorithms available in WEKA with our dataset using cross-validation with 5 fold and 10 fold with test options available.

Finally we found that cross-validation give the best classification using 10 Fold cross-validation with the reduced the dimensionality obtained in the previous step to six attributes selection . Then we selected algorithms with classification accuracy between 90% to 100% as the proposed Base Classifiers. The Base algorithms Proposed in our investigation in this research are: K- nearest neighbor (IBk), Attribute Selected algorithm, Bagging, Random Committee, Rule-based learning (PART), Decision tree algorithm J48, Logistic Model Trees(LMT), Random Forest, Random Tree, as illustrated in Table 5.2 the correctly classified for each base classifier in term of percentile using 5 and 10 fold cross-validation.

Table 5.2: Base classifiers proposed

6 Selection Attribute		
Test Options: Cross-Validation		
	5 Fold	**10 Folds**
Classifiers	**Correctly Classified**	**Correctly Classified**
IBk	88.7248 %	90.3356 %
Attribute Selected Classifier	89.1275 %	91.9463 %
Bagging	88.4564 %	90.4698 %
RandomCommittee	93.557 %	95.0336 %
PART	89.9329 %	91.8121 %
J48	89.6644 %	92.8859 %
LMT	92.6174 %	92.2148 %
Random Forest	92.8859 %	94.2282 %
Random Tree	92.8859 %	94.8993 %

Table 5.3 depicts the time required to build the model for each algorithm. From Table 5.3 it is inferred that Random Tree model and IBK classifies are very quickly in comparison to other models.

Table 5.3: Time required building the model

Algorithm	**Time taken to build model (in seconds)**
IBk	0.001
Attribute Selected Classifier	0.14
Bagging	0.12
Random Committee	0.05
PART	0.04
J48	0.02
LMT	2.46
Random Forest	0.04
Random Tree	0.001

Table 5.4 depicts the various error metrics analyzed in the data set. It is inferred from Table 5.4, that Random Committee has the highest Kappa Statistic value and also has better accuracy compared with the others classifiers. Hence Random Committee is an appropriate model for classifying the hospital situation with higher accuracy.

Table 5.4: Performance measures comparison

Algorithm	MAE	RMSE	KS	Correctly Classified
IBk	0.0978	0.3104	0.8062	673 90.3356 %
Attribute Selected Classifier	0.1008	0.2631	0.8384	685 91.9463 %
Bagging	0.1527	0.2609	0.8089	674 90.4698 %
Random Committee	0.0643	0.1931	0.9004	708 95.0336 %
PART	0.101	0.264	0.8355	684 91.8121 %
J48	0.0865	0.2518	0.8574	692 92.8859 %
LMT	0.0854	0.2454	0.844	687 92.2148 %
Random Forest	0.0961	0.219	0.8843	702 94.2282 %
Random Tree	0.051	0.2258	0.8977	707 94.8993 %

In the next step we investigated and analyzed the performance of the Base Classifiers Proposed obtained in the previous step using the same wearable sensors dataset with the reduced number of attributes.

The investigation and analysis of the performance is based on:

- Classifier performance in term of recall precision, f measure and false alarm rate
- Classification performance for normal class.
- Classification performance for abnormal class.
- Classification performance of each classifier in term of Sensitivity and Specificity.

The following Tables and Figures illustrate our experimental results with various evaluation methods. Table 5.5 depicts the classifier performance of each classifier in term of recall precision, *f* measure and false alarm rate. It is inferred from Table 5.5 that Random Committee model has the highest precision and lowest false alarm rate, and the same recall as Radom Tree.

Table 5.5: Performance in term of recall precision, *f* measure and false rate

Algorithm	**Recall**	**Precision**	**F-measure**	**False alarm rate**
IBk	0.916	0.905	0.911	0.085
Attribute Selected classifier	0.914	0.914	0.913	0.076
Bagging	0.893	0.905	0.898	0.084
Random Committee	0.938	0.957	0.947	0.038
PART	0.924	0.908	0.914	0.080
J48	0.9185	0.9316	0.924	0.061
LMT	0.905	0.9316	0.917	0.062
Random Forest	0.927	0.951	0.938	0.044
Random Tree	0.9383	0.9544	0.945	0.041

As an example of classifier error illustration, Figure 5.2 depicts the Classifier error of Random Committee. The blue crosses indicate the Normal class and red crosses indicate the Normal class and squares indicate not classified.

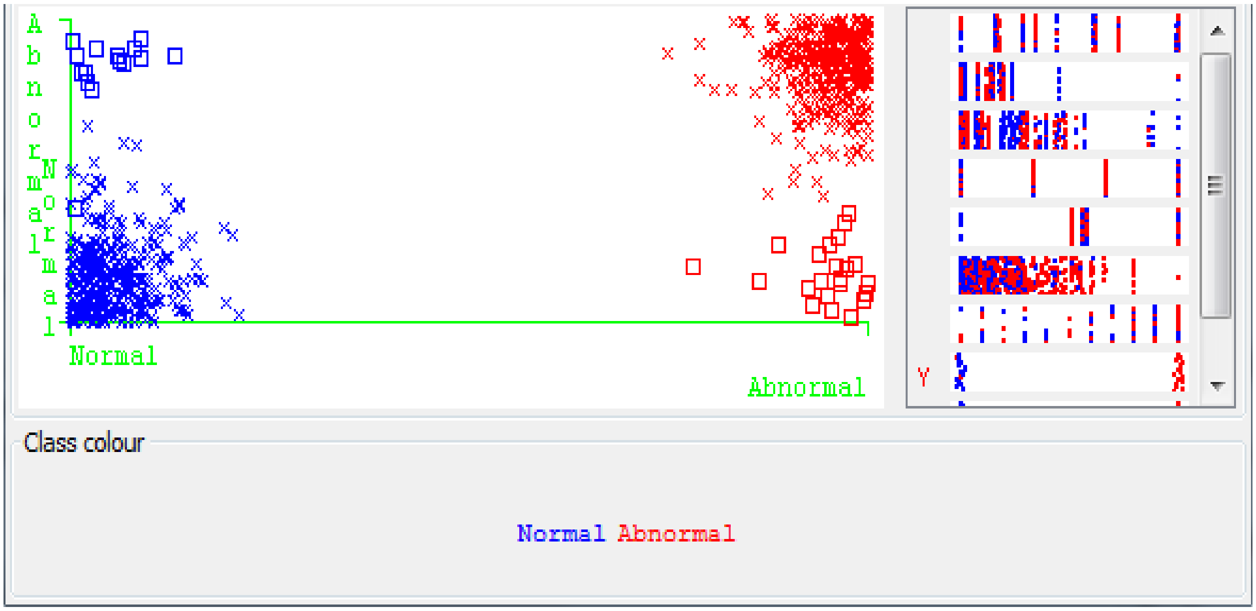


Figure 5.2: Classifier error of random committee

Table 5.6 depicts the algorithm performance of each classifier in term of recall precision and *f* measure for Normal class is summarized. It is inferred from Table 5.6 that Random Committee model has the highest precision and also high recall.

Table 5.6: Classification performance for normal class

Classifiers	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
IBk	0.883	0.109	0.878	0.883	0.881	0.891
Attribute selected classifier	0.906	0.122	0.869	0.906	0.887	0.926
Bagging	0.880	0.112	0.875	0.880	0.878	0.953
Random Committee	0.943	0.071	0.922	0.943	0.932	0.984
PART	0.926	0.124	0.869	0.926	0.897	0.955
J48	0.903	0.109	0.881	0.903	0.892	0.934
LMT	0.886	0.094	0.894	0.886	0.890	0.948
Random Forest	0.937	0.084	0.909	0.937	0.923	0.973
Random Tree	0.932	0.074	0.919	0.932	0.925	0.929

Figure 5.3 depicts the Area under ROC of Random Committee classifier with highest area under Roc.

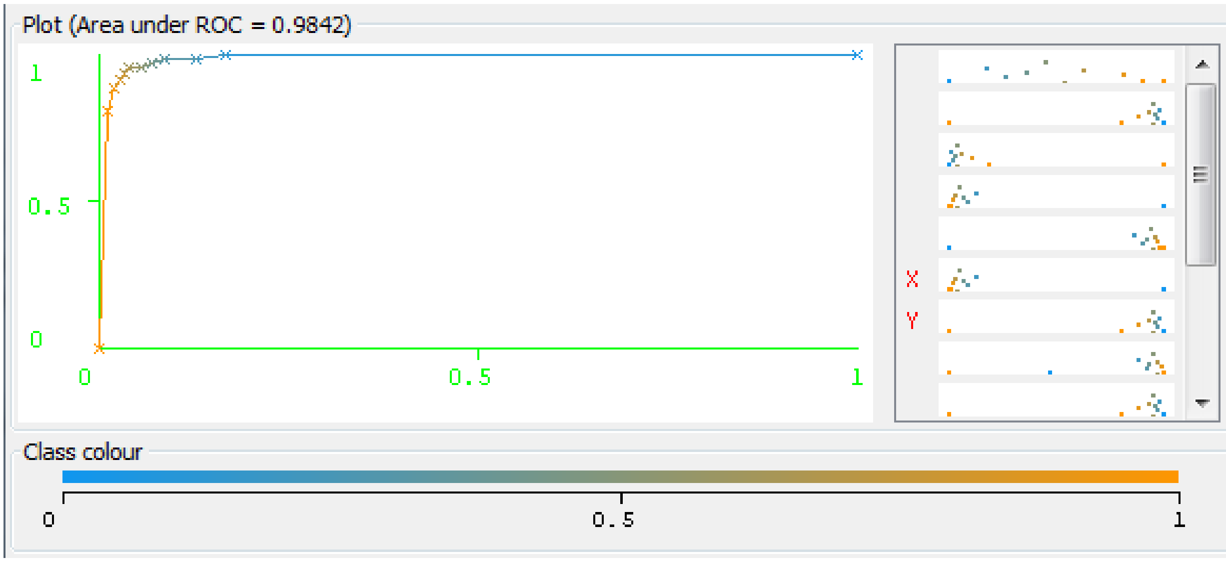


Figure 5.3: AUR of random committee classifier

Table 5.7 depicts the classifier performance of each classifier in term of recall precision, and f measure for abnormal class. It is inferred from Table 5.7 that Random Committee model has the highest precision.

Table 5.7: Classification performance of each classifier for abnormal class

Classifiers	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
IBk	0.901	0.094	0.915	0.901	0.908	0.910
Attribute selected classifier	0.878	0.094	0.913	0.878	0.895	0.926
Bagging	0.888	0.120	0.893	0.888	0.891	0.953
Random Committee	0.929	0.057	0.948	0.929	0.938	0.984
PART	0.876	0.074	0.930	0.876	0.902	0.955
J48	0.891	0.097	0.912	0.891	0.901	0.934
LMT	0.906	0.114	0.899	0.906	0.903	0.948
Random Forest	0.916	0.063	0.943	0.916	0.929	0.973
Random Tree	0.926	0.068	0.938	0.926	0.932	0.929

Table 5.8 depicts the classification performance of each classifier in term of Sensitivity and Specificity with the Random Committee model having the highest Specificity and also high

Sensitivity. Random Committee model also has the highest accuracy and the IBK model has the lowest accuracy.

Table 5.8: Performance model in term of sensitivity and specificity.

Classifiers	Sensitivity	Specificity	Accuracy
IBk	0.8907	0.914	0.9033
Attribute Selected Classifier	0.941	0.923	0.9194
Bagging	0.8932	0.915	0.9046
Random Committee	0.938	0.961	0.9503
PART	0.924	0.92	0.9221
J48	0.918	0.938	0.9288
LMT	0.905	0.937	0.9221
Random Forest	0.927	0.955	0.9422
Random Tree	0.938	0.958	0.9489

Table 5.9 depicts overall algorithm performance ranked by accuracy. It inferred from Table 5.9 that Random Committee has the highest accuracy and the IBK model has the lowest accuracy comparing with the rest of the classifiers in this research.

Table 5.9: Overall algorithm performance ranked by accuracy

Algorithm	Accuracy
Random Committee	0.9503
Random Tree	0.9489
Random Forest	0.9422
J48	0.9288
PART	0.9221
LMT	0.9221
Attribute Selected Classifier	0.9194
Bagging	0.9046
IBk	0.9033

5.5 Second Phase Experiments and Results

5.5.1 Novel Ensemble Decision Support and Health Care Monitoring System

As stated in methodology section and according to literature, in this thesis the main purpose of an ensemble methodology is to combine a set of models, each of which solves the same original problem, in order to obtain a better composite global model with more accurate and reliable estimates or decisions than can be obtained from using a single model.

This Section is devoted to extensive investigation to construct a new novel ensemble healthcare decision support for assisting an intelligent health monitoring system. Experiments are conducted using the same dataset with the reduced dimensionality of the attributes obtained in the phase one experiments.

This second phase consist of two steps. First step, extensive investigation of the experimental results of the performance of different Meta classifiers techniques (see Annex B**)** for classifying the dataset. Second step, in this second phase experiments the researcher selected five of the classifiers with highest performance accuracy obtained in the first phase experiments (see Table 5.9) as base classifiers in the second phase experiment. These base classifiers used in this phase: Random Tree, Random Forest, J48, PART, LMT to construct a novel ensemble model. Comparative analysis and evaluation have been done using various evaluation methods and the performance factors used for analysis are accuracy and error measures. The accuracy measures are TP rate, F Measure, ROC area, Sensitivity and Specificity. The error measures are Mean Absolute Error, Root Mean Squared Error and Kappa Statistics.

First step, we tested various Meta classifiers and have chosen the following classifiers for a series of complete tests with outcomes presented in this research. We found that cross-validation give the best classification with 10 Fold. These Meta classifiers used are AdaBoostM1, Bagging, LogitBoost, Random Committee, Stacking, and Voting as depicted in Table 5.10.

Table 5.10 depicts the various error metrics analyzed in the data set. It is inferred from Table 5.10 that Random Committee has the lowest MAE and highest Kappa Statistic value, Random Committee is an appropriate model for classifying the hospital situation with MAE = 0.06 and 95.0336 % were correctly classified.

Table 5.10: Performance measures comparison of individual Meta classifiers

Meta - Classifier	MAE	RMSE	KS	Correctly classified
AdaBoostM1	0.2957	0.3794	0.6051	599 80.4027 %
Bagging	0.1527	0.2609	0.8089	674 90.4698 %
Logit Boost	0.2725	0.3593	0.644	613 82.2819 %
Random Committee	0.0643	0.1931	0.9004	708 95.0336 %
Stacking	0.4983	0.4992	0	394 52.8859 %
Vote	0.4983	0.4992	0	394 52.8859 %

Second step, the researcher investigated and constructed various Voting ensembles by combining methods based on meta classifier Voting and combined with previous selected base classifiers obtained in phase one in previous subsection.

Table 5.11 depicts various ensemble models of Meta Voting Classifiers combining with various single classifiers. Voting combining: J48, LMT, Random Forest, Random Tree, PART (Voting + 5 classifiers), Voting combining: J48, Random Forest, Random Tree (Voting + 3 classifrs), and Voting combining: Random Forest, Random Tree (Voting + 2 classifiers)

Table 5.11. Combined classifiers

Combined Classifiers	Base Classifiers				
	J48	**LMT**	**Random Forest,**	**Random Tree**	**PART**
Voting + 5 classifiers	J48	LMT	Random Forest	Random Tree	PART
Voting + 3 classifiers	J48,	-----	Random Forest	Random Tree	----
Voting + 2 classifiers	------	------	Random Forest	Random Tree	-----

Tables 5.12 depicts the classifier performance of each classifier in term of MAE, RMSE, Kappa statistic, Time to build a model and % Correctly Classified. It is inferred from Table 5.12 that the ensemble (Voting + 3 classifiers) has the least MAE and RMES than ensemble (Voting +

5 classifiers) and the same Kappa Statistic value as (Voting + 5 classifiers). Ensemble (Voting + 3 classifiers) has the highest MAE and RMES than ensemble (Voting + 2 classifiers) and the highest Kappa Statistic value than (Voting + 2) classifiers. But in terms of Correctly Classified instances, ensemble (Voting + 3 classifiers) has the highest Correctly Classified instances than the other considered ensembles.

Table 5.12: Performance measures comparison for Ensemble models

Combined Classifiers	Correctly Classified	MAE	RMSE	Kappa statistic	Time to build a model
Voting + 5 classifiers	710 95.302 %	0.1239	0.2206	0.906	2.51 seconds
Voting + 3 classifiers	711 95.436 %	0.1025	0.2077	0.9086	0.07 seconds
Voting + 2 classifiers	707 94.899 %	0.0866	0.204	0.8977	0.05 seconds

It is inferred from Tables 5.12 and 5.14 that ensemble (Voting + 3 classifiers) has the best Correctly Classified than all individual Meta Classifiers, but individual Meta Classifiers Random Committee has the lowest MAE and RMSE than the Ensembles combined model.

Table 5.13 depicts the performance of each classifier in term of recall precision and f-measure and false alarm rate. It is inferred from table 13 that Ensemble (Voting + 3 classifiers) model has the highest F-measure and the highest precision as (Voting + 5 classifier) and lowest false alarm rate. In the term of recall with the same recall value as (Voting + 2 classifiers), and highest recall value than (Voting + 5 classifiers).

Table 5.13: Performance in term of recall precision, and false rate

Ensemble	Recall	Precision	F-measure	False Alarm rate
Voting + 5 classifiers	0.9270	0.97720	0.951433	0.02133
Voting + 3 classifiers	0.9318	0.9743	0.95257	0.023809
Voting + 2 classifiers	0.9383	0.9544	0.94627	0.041237

Table 5.14 depicts the algorithm performance of each classifier in term of recall precision and f- measure for Normal and Abnormal classes is summarized. It is inferred from Table 5.14 that (Voting + 5 classifiers) model has the highest ROC Area and also highest PRC Area than the

others Ensemble in classification the class Normal and Abnormal classes but in term of F-Measure the ensemble (Voting + 3 classifiers) has highest F- Measure the others.

Table 5.14: Performance classification of combining models in term of ROC .

Ensemble	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	PRC Area	Class
Voting + 5 classifiers	0.977	0.069	0.927	0.977	0.951	0.987	0.985	Normal
	0.931	0.023	0.979	0.931	0.954	0.987	0.986	Abnormal
Voting + 3 classifiers	0.974	0.063	0.932	0.974	0.953	0.982	0.972	Normal
	0.937	0.026	0.976	0.937	0.956	0.982	0.979	Abnormal
Voting + 2 classifiers	0.954	0.056	0.938	0.954	0.946	0.981	0.969	Normal
	0.944	0.046	0.959	0.944	0.951	0.981	0.981	Abnormal

In term of sensitivity, specificity. Table 5.15 depicts the classification performance of each classifier in term of Sensitivity, Specificity. It is inferred from Table 5.15 that the ensemble (Voting + 3 classifiers) model has the highest Accuracy than the others ensemble. But in terms of specificity and sensitivity the Ensemble (Voting + 2 classifiers) is highest.

Table 5.15: performance combining modeles in term of sensitivity and specificity

Combined Classifiers	Parameter	Sensitivity	Specificity
Voting + 5 classifiers		0.92702	0.93147
Voting + 3 classifiers		0.93188	0.93654
Voting + 2 classifiers		0.93837	0.94416

Table 5.16 depicts the overall Ensemble Voting performance ranked by accuracy. It is inferred from Table 5.16 that Ensemble (Voting + 3 classifiers) model has the highest accuracy and the Ensemble (Voting + 2 classifiers) model has the lowest accuracy.

Table 5.16: Overall ensembles performance ranked by accuracy

Algorithm	Accuracy

Voting + 3 classifiers	0.95436
Voting + 5 classifiers	0.95302
Voting + 2 classifiers	0.94899

Table 5.17 depicts the overall Ensembles and Meta classifiers performance ranked by accuracy. It is inferred from Table 5.17 that Ensembles (Voting + 3 classifiers) and (Voting + 3 classifiers) models have the highest accuracy and the Meta classifiers Stacking and Voting models have the lowest accuracy.

Voting provided better results compared to individual classifiers and meta classifiers. Ensemble (Voting + 5 classifiers) or ensemble (Voting + 3 classifiers) yielded similar results.

Table 5.17: Ensembles and meta classifiers performance ranked by accuracy

Models	**Accuracy**
Ensemble (Voting+ 3 classifiers)	0.95436
Ensemble (Voting + 5 classifiers)	0.95302
Random Committee	0.95033
Ensemble (Voting + 2 classifiers)	0.94899
Bagging	0.90469
Logit Boost	0.82281
AdaBoostM1	0.80402
Stacking	0.52885
Vote	0.52885

Figure 5.4 depicts the classification error of Ensemble (Voting+ 3 classifiers) performance, the blue crosses indicated Normal class classification, the red crosses indicate the Abnormal class classified, the red squares indicated Abnormal class unclassified and the blue squares indicated Normal class unclassified.

Figure 5.4: Error of ensemble (voting+ 3 classifiers) performance

Figure 5.5 depicts the abnormal Class, Area under Roc of Ensemble (Voting + 3 classifiers) with highest area under ROC (0.9816).

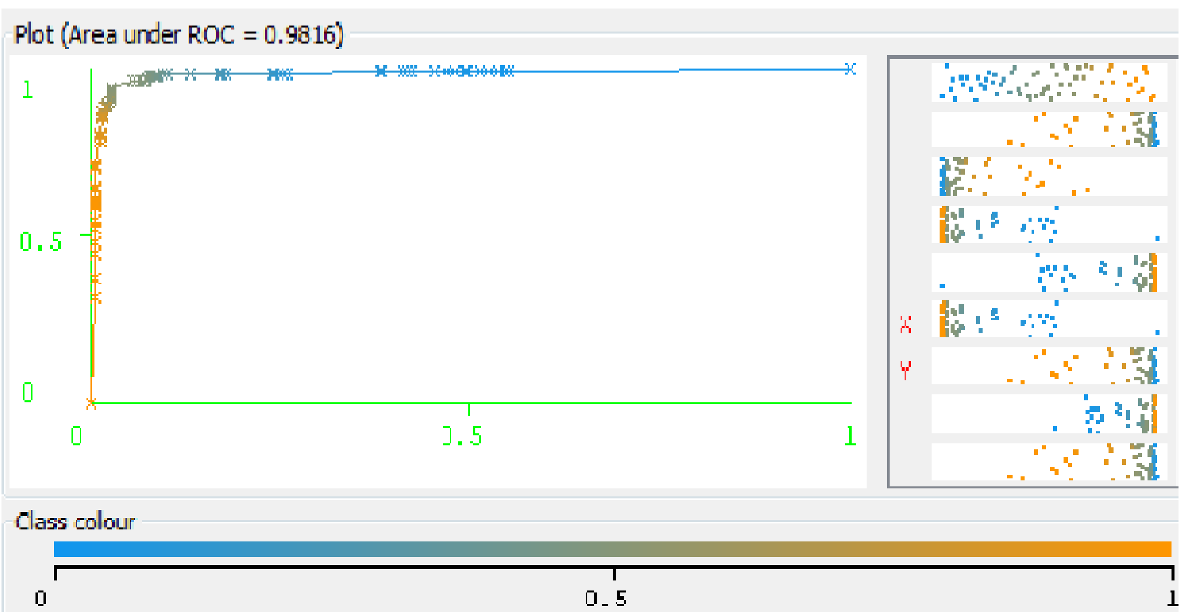


Figure 5.5: Abnormal class, area under ROC of ensemble (voting+ 3 classifiers).

Figure 5.6 depicts the results when the cost is 0, Random is 394 and the difference between the values of the cost function between the random selection and the current value of the cost is called Gain, as indicated in the right side of the frame. In the context of abnormal situation, the Gain can be interpreted as the benefit obtained by using the classification model instead of random selection of the same number of patients.

In our experiments, the gain (Benefit) obtained is 0. Threshold curve depicts the dependence of the part of class "Abnormal" patients retrieved in the course of predicting selected from the whole dataset (i.e. only those selected for which the estimated probability of having abnormal disease exceeds the chosen threshold). The confusion matrix for the current value of the threshold is shown in the Confusion Matrix frame at the left bottom corner of the window.

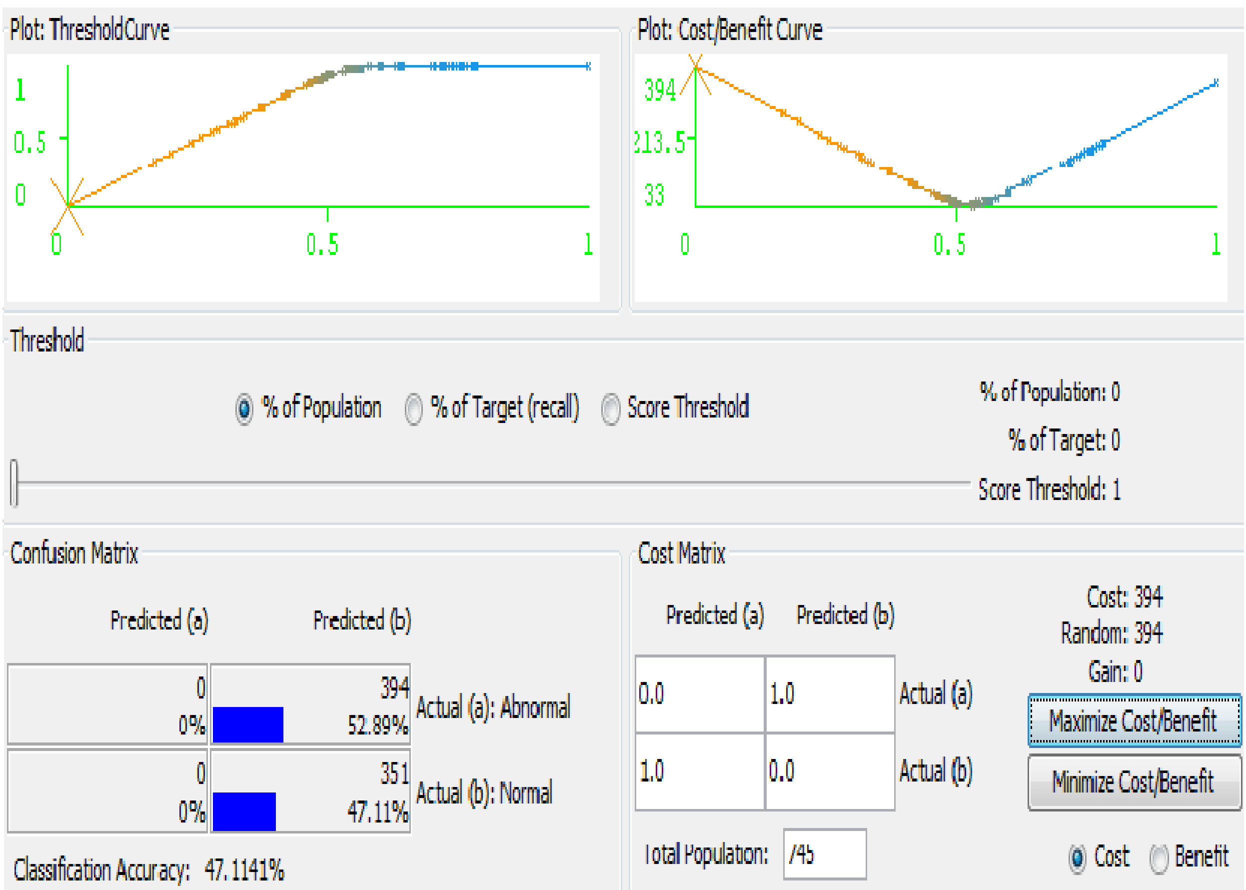


Figure 5.6: Maximize cost/benefit of class abnormal

Figure 5.7 depicts the results when the cost is 33, Random is 370.97 and the Gain is 337.97. In the context of abnormal disease, the Gain can be interpreted as the benefit obtained by using the classification model instead of random selection of the same number of patients. In our experiments, the gain (Benefit) obtained is 337.97 and the classification Accuracy is 95.5705, which means that using Cost/Benefit, we can obtain more classification accuracy than ROC Curve.

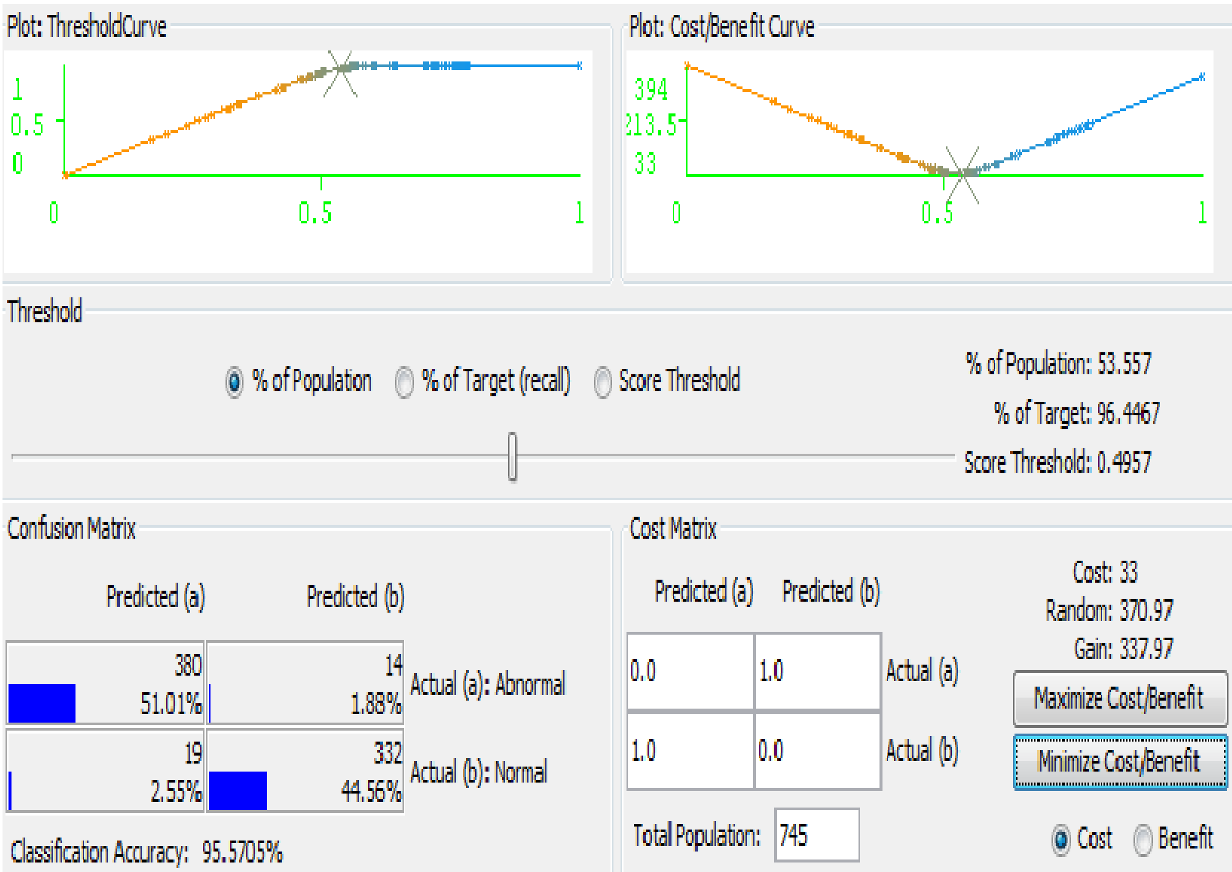


Figure 5.7: Minimize cost/benefit of class abnormal

5.6. Models Comparison, Discussion and Lessons Learned

As mentioned in previous Sections, two major phases of experiments were conducted. In the first phase experiment, attempt has been made to investigate the experimental results of the performance of different classification techniques for classifying the data from different simulated wearable sensors used for monitoring different patients with different diseases. We explored and evaluated the models with various methods of evaluation based on Error Metrics, ROC curves, Confusion Matrix, Sensitivity and Specificity.

Empirical results indicate that the execution time of Random Committee algorithm is lowest for classification in comparison with the rest of classification algorithms, and the LMT algorithm has the higher execution time. The MSE error of the classification values for Random Committee is lower in comparison with the rest of the based proposed classifiers, and the Meta bagging classifier has higher MSE error in comparison with the rest of the base proposed classifiers. In terms of recall precision, f measure and false alarm rate the Random Committee model has the highest precision and lowest false alarm rate, and the same recall as Random Tree.

In term of recall precision and f- measure for Normal class, it is inferred that Random Committee model has the highest precision and also high recall. With higher true positive rate and minimum false rate also with higher ROC Area when the classification is Normal class in comparison of the rest of the classifiers. Attribute Selected Classifier has the lower precision in comparison with the rest. Also from the performance of each classifier in term of recall precision and f measure for abnormal class, Random Committee model has the highest precision and also high recall (with higher true positive rate and minimum false rate), also has highest ROC Area in comparison with other classifiers. While PART classifier has the lowest precision the same as Attribute Selected Classifier but with highest in ROC Area compare with Attribute Selected Classifier.

From Sensitivity, Specificity and Accuracy perspective, the Random Committee model has the highest Specificity and also high Sensitivity the same as Random Tree but with highest accuracy of all the classifiers. While IBK classifier has the lowest Sensitivity, Specificity and Accuracy compare with the rest of the classifiers.

To sum up, from the execution and accuracy point of view, Random committee model can be identified as the best choice for analysis and decision model among all the other classifier algorithms. Random committee provides an advantage that with a reduced attribute set a better classification performance. Empirical results illustrate that Random committee classifier with selection attribute method gives better accuracy, error rate and reduced false alarm rate and with the highest Sensitivity and Specificity.

The aim of the second phase, is to explore various ensembles combining models and evaluate the models with various methods of evaluation based on Error Metrics, ROC curves, Confusion Matrix, Sensitivity, Specificity and the Cost/Benefit methods. To construct a new novel ensemble healthcare decision support for assisting an intelligent health monitoring system. Experiments are conducted using the same dataset with the reduced dimensionality of the attributes obtained in the phase one experiments.

We summarize the obtained results from the evaluation conducted in the previous Sections. The results indicate that the execution time of Ensemble (Voting + 2) classifiers algorithm is lowest for classification in comparison with the rest of ensemble classification algorithms, and the Ensemble (Voting + 5 classifiers) classification algorithm has the higher execution time.

The MSE error of the classification values for Ensemble (Voting + 2 classifiers) is lower in comparison with the rest of the based proposed classifiers, and the Ensemble (Voting + 5 classifiers) classifier has higher MSE error in comparison with the rest of the base proposed classifiers. In terms of recall precision, *f* measure and false alarm rate the Ensemble (Voting + 5 classifiers) model has the highest precision and lowest false alarm rate, and the has the highest recall lower in comparison with the rest of the ensembles models.

In term of recall precision and f- measure for Normal class it is inferred that Ensemble Voting + 3 classifiers model has the highest precision than Ensemble Voting + 5 classifiers model, but with lowest recall than Ensemble (Voting + 5 classifiers), has highest recall and highest TP Rate than Ensemble (Voting + 2 classifiers), and with minimum false rate than Ensemble (Voting + 5 classifiers) also with higher Roc Area and higher PRC when the classification is Normal class in comparison of the Ensemble (Voting + 2 classifiers). The Ensemble (Voting + 5 classifiers) has higher ROC Area and higher PRC when the classification is Normal class in comparison with the rest. In the case of class Abnormal we found that Ensemble Voting + 3 classifiers has highest True Positive Rate, minimum false rate and highest recall in comparison with the rest.

We found the Ensemble (Voting + 5 classifiers), has highest ROC Area and higher PRC when the classification is abnormal class in comparison with the rest. From Sensitivity, Specificity and Accuracy perspective, the Ensemble Voting + 2 classifiers model has the highest Specificity and also high Sensitivity followed by Ensemble (Voting + 3 classifiers) model. From Accuracy perspective, the Ensemble (Voting + 3 classifiers) model has the highest Accuracy in comparison with the rest.

To sum up, from the execution and accuracy point of view, Ensemble (Voting + 3 classifiers) or Ensemble (Voting + 5 classifiers) model can be identified as the best choice for analysis and detection model among all the other classifier ensembles modes algorithms for our data set. Ensemble (Voting + 3 classifiers) or Ensemble (Voting + 5 classifiers) provides an advantage that with a reduced feature set a better classification performance and is able to offer a better decision support system.

The last evaluation method used in our experiments is Cost/Benefit method. As indicate in the result section using Cost/Benefit method minimizes the cost and increases the classification accuracy. In our experiment the gain (Benefit) obtained is 337.97 and the classification Accuracy is 95.5705, which means that using Cost/Benefit we can obtain more classification accuracy than ROC Curve.

The main Goal of this Section is to evaluate ensemble design and combining different algorithms to develop a novel intelligent ensemble healthcare decision support and a monitoring system to classify the situation of an emergency hospital based on the Vital Signs from Wearable Sensors. We compared the performance of the entire Individual base classifiers, Meta classifiers and Ensemble combine models. Empirical results illustrate that Voting combining with J48, Random Forest, Random Tree (Voting + 3 classifiers) model, or J48, LMT, Random Forest, Random Tree, PART, with selection attribute method gives better accuracy, with high recall and high f- measure. Our Novel Intelligent Ensemble Health Care Decision Support and Monitoring can optimize the results and improve assisted health care monitoring.

5.7 Data Analysis Healthcare Monitoring Trends and its Effects

As discussed in Literature Review, there were research attempts on healthcare monitoring data analysis in the past few years. A look at data mining research reveals that healthcare monitoring analysis from patients monitoring, data occurrence and severity perspective requires due attention as variations in results are exhibited. Partly, variation is due to the techniques and data input used.

In addressing the second main objective, decision support models using ensemble classifiers through voting technique combining with J48, Random Forest, Random Tree, are found to be effective in analysis of healthcare monitoring data. The attributes selection method is used to reduce an input to the content dimension of the healthcare monitoring information architecture. Magnitude of the data quality problems was also addressed to empirically show the extent of the problem and the need for architectural guideline in managing patients monitoring information.

The insight gained so far in the process of the research, revealed that healthcare monitoring data should be complete for the analysis to reflect important patterns and decision support. There should also be a mechanism for data quality checks about each patient monitoring data that requires architectural guideline. Once data is organized, machine learning approaches to real time analysis of the patients monitoring data is required to see changes through time and adjust the counter measures accordingly.

These findings are reflected on the information architecture proposed as a guideline for healthcare monitoring data collection and analysis. The design of the architectural guideline will create a platform for the analysis of **AmIHCM** in supporting efforts to improve healthcare monitoring.

This research perceives healthcare monitoring data collection and analysis as a system that requires a special view towards understanding the whole and making sense out of it for improved decision support in the effort of reducing the problem of healthcare monitoring ultimately. That is why the issue of data quality and understanding gets attention in addition to decision support modeling.

5.8 Analysis of Healthcare Monitoring Information Management in Khartoum State Hospitals

5.8.1 Introduction

This sub-section presents findings of qualitative data analysis into the healthcare monitoring information management. Accordingly, one of the research sub questions of this thesis, mentioned below, is addressed through the analysis of qualitative data collected.

RQ3. What are the problems of the current patients vital signs monitoring data collection, analysis and dissemination practice on the **HCM**?

In addressing this research question we define two sub objectives as presented below:

- To examine problems related to patients vital signs monitoring data reporting, data quality and analysis mechanisms in a **HCM** domain.
- To identify the structure and requirement of **AmIHCM** data collection and analysis focusing on healthcare vital signs monitoring data.

5.8.2 Data Collection and Analysis Process

To deal with the specific research objectives mentioned, this research involved was conducted at Khartoum state hospitals departments operating on patient’s vital signs monitoring data collection. Accordingly, healthcare monitoring experts in Khartoum State hospitals, metros and nurses in hospitals, and statistical in statistic departments at hospitals who are considered to be central in the overall healthcare monitoring systems were interviewed.

According to , sampling method is the most frequently used method in qualitative research. In this research sampling approach was adopted. There is no general agreement among researchers about sample size in qualitative research as the data collection continues until reasonable saturation. Literature in the qualitative research area suggests sample sizes ranging from 3 to 30. suggests minimum sample size less than or equal to 10 interviews. 15 to 20 participants were suggested by and 20 to 30 participant by for Grounded Theory . However, as stated by sample sizes should not also be too small for it is difficult to achieve theoretical and / or data saturation.

In this research, the sampling was based on healthcare monitoring in twenty-five selected hospitals in Khartoum state (Sudan). Seven private hospitals and eighteen public hospitals also at the ministry of health in Khartoum State, total of 25 participants. This makes the research more comprehensive and representative; the majority of the registered patients are concentrated at the capital Khartoum State. The sampling frame comprises all stakeholder organizations working on healthcare monitoring data, which includes experts, metros, Nurses in Khartoum state hospitals, statisticians in statistics department at hospitals and in the Ministry of health in Khartoum state.

The design of the data collection protocol was divided into 3 tasks: preparing interview questions based on **ZF** dimensions, determine the subject to be interviewed and initial scheduling of field visit. The data collection protocol was constructed to ensure consistency across multiple discussions and interviews. The research followed the **ZF** as a theoretical base in a structured way for acquiring the necessary knowledge about healthcare monitoring Khartoum State hospitals practice with respect to the Patients monitoring data information management. Thus, in the process of collecting data for this research and in order to answer the questions triggered by the intersecting factors (cells) from the top three rows and six columns of **ZF**, qualitative data-gathering techniques mentioned in the methodology section were used.

Accordingly, the interview questions (see Annex C and Annex D) were designed in such a way that they enable to collect relevant facts regarding the three aspects of the research; Patients monitoring data collection and reporting, data handling and information sharing, data analysis and dissemination from the six dimensions of the **ZF** namely**; why, what, how, who, when** and **where.** Questions asked probed data content and process requirements, people, network and time of the process, difficulties faced by the healthcare monitoring experts and statistical data analysts, and the supporting architecture for healthcare monitoring data collection and analysis.

The interview questions were open ended to gather as detailed information as possible. The questions were categorized into general and specific. The general questions were about the magnitude of the problem in general and practices related to patients monitoring information management. In the second category, six issues related to the six dimensions of the Zachman framework were covered through a set of motivational questions. Accordingly, the first set of questions was related to the structure /content /data **“what”** requirement of the domain. The second set of questions guided the discussion about the required functions and process, **“how”**, of the domain under study. The third set of questions was about the prime motivation, **“why”**, of patients monitoring data management. In the same line, the fourth set of questions focused on the geographical location **“where”** of patients monitoring data reporting, analysis and dissemination while the fifth set of questions were about the timing **“when”** of the processes and services in the domain under consideration. The last set of questions were related to the stakeholders or people **“who”** in the patients monitoring information management.

In the process of data collection, the researcher first explained the purpose of the research and got the respondents’ agreement to participate in the research. The researcher interviewed all healthcare monitoring experts, for between 45 and 60 minutes. All interview responses were noted, with the consent of the respondents. After the interviews, the short notes were transcribed into computer files. Care was taken to consider all ethical issues. The interview focused primarily, on issues related with patients monitoring data reporting, analysis and dissemination (see Annex C).

As indicated in the methodology section, interview results are also supported by repeated observations. The purpose of the observation was to best understand the central theme of the research, as stated it in . In addition to repeated observations, relevant documents were collected. The documents were collected up on request from the hospitals managers. These official documents were important to the study as they deal with patients monitoring vital signs information related issues.

In qualitative research literature, analysis of the data starts right from the first day of data collection. It basically, involves major steps like preparing and organizing data for analysis, reading through all data, arranging in themes along with description, interrelating theme’s description and interpreting them. These activities can be categorized in to three namely; data reduction, data display and conclusion drawing. Regarding data reduction attempt has been made to transform the raw data from all data source in to a table format based on the interview questions and themes identified. A portion of it is attached as Annex E. With respect to data display narrative presentation along with supporting quotations was followed as shown in this section.

Conclusion is then drawn to design the **AmI** healthcare monitoring Information Architecture. As indicated above, in applying **ZF**, and to holistically control the study, a qualitative case study approach with multiple data collection techniques mentioned in the previous sections were adopted. Information obtained through these data gathering techniques helps to populate the requirements of the top three layers (18 cells) in **ZF** to ascertain the design details of the healthcare monitoring information management context, business and system model.

The framework abstracts the characteristics and features of the healthcare monitoring information based on six dimensions, **Motivation, Data, People, Process, Place and Time,** and explain their structures and processes from the perspectives of the **planner, owner and designer** of the healthcare monitoring information.

Data analysis was carried out using thematic coding methods . The thematic coding is based on the dimensions the selected framework. Findings were then used to populate the Zachman Framework at the top three layers, namely the scope, business model and the system model. Particularly the analysis results are used to define the six dimensions of the first row in the Zachman Framework, which will be then used to derive the lower level cell contents. All the interview transcripts were read by the researcher and coded in the style of a grounded theory approach to data analysis . Our research identified a number of patient monitoring data, process, people, motivation, place and time issues that influence healthcare information management. Three categories were generated and under these all of the data were represented. The reporting style followed is an interpretive qualitative research reporting model where a summary of findings are presented followed by illustrative quotes of respondents and interpretations of the researcher. Findings from the data collection process are presented in three themes.

With the overall objective of the research project in mind, to define an IA that forms the basis of AmI healthcare monitoring information management system, the general observation from the data collection process revealed that there are definitely shortcomings in patients monitoring data collection, analysis and dissemination at various levels requiring an immediate attention and architectural guideline. The interview process, as well as observation made during group discussion with the various stakeholders within the healthcare monitoring system, gave the researcher the necessary confirmation of the problems and information requirement of the domain. The detail is presented below in three themes preceded by a general issues regarding patient's healthcare monitoring information management practice.

The interview processes were started with two general questions: How do you explain the healthcare monitoring situation in Khartoum state hospitals? How do you describe the effectiveness of patient's monitoring data collection and analysis practice?

In responding to the first question, it was easy to learn the respondent's agreement on the magnitude of the healthcare monitoring problem. They explained the situation as a very problematic issue and top in the priority list. Apart from investigation and enacting legal actions on the accused party, lose attention and disintegrated efforts were cited as major problems in the effort of healthcare monitoring problem reductions. One respondent commented that: since healthcare monitoring department mainly focuses on only monitoring patients vital signs, analysis of data processing is absent. This fact was also observed by the researcher.

5.8.3 Content and Process of healthcare monitoring Information Management

The first question in relation to the content and process of healthcare monitoring information management was regarding patient monitoring data sources for which all respondents confirmed the fact that there was no integration of data from the different sources. **HCM** department record was mentioned as the only data sources. One respondent expressed his view that the data incompleteness is very serious problem at a hospital. In responding to another question, general concerns regarding the integrity of completeness and usefulness of healthcare monitoring information provided have been expressed by the interviewees.

This is exhibited by one interviewee as "There are many cases where healthcare monitoring department face the difficulties to communicate with doctor or relatives when there are chronic cases".

Another question in the interview process allowed respondents to comment on the data content about patient monitoring. A number of these responses outlined data items that healthcare monitoring data system should contain as analysis of patients monitoring data, real time report, and integration with others departments, Doctor, Nurse Schilling, relative's data and environment. These concepts were described in a form of six basic questions as where, when, why/how patient monitoring data occurs; who was involved, what are the environmental conditions and the result of a conflict, which patient monitoring data system should respond.

Regarding the process of patient's healthcare monitoring data collection and analysis respondents were asked to describe the patient monitoring data reporting and analysis process.

Accordingly, it is learned that the collection of patients monitoring data in Khartoum State Hospitals is the responsibility of the healthcare patient's vital signs monitoring department in hospitals and there are absent of analysis of patients monitoring data and real time report.

The process starts immediately as the patients arrives to hospital with chronic disease after primary healthcare was done, and depending on the status of the patient, room will assigned to each patient and a Nurse will be responsible for measurement of the patient vital signs. The Nurse records the vital signs measurement manually in the vital signs sheet, every three hours or some time the measurement depends on the status of the patient. The nurse will present this vital signs sheet to the doctor when he comes to see the patient. The doctor will write the aspirate treatment to each patient according to the vital signs sheet, patient diagnosis and the status of the patient.

Though there are attempts to develop a standard patient vital signs recording format at a hospital level, currently patient monitoring data collection is in a long hand sheet written format without daily recording in a computerize file and database. With the exception in one hospital, where at least measurement patient's vital signs resulting are recorded in an excel format file. Despite the relevance of different information at the patient's monitoring scene, not all the data and information is recorded, resulting in incomplete and also in accurate data. After the patient leaves the hospital or dead the patient vital signs sheet form will be send to the statistical department so as data could be archived.

However, by the time of this research it can be said that almost all the data handling and storage is being done on a sheet form based on a daily patient measurement vital signs form, without recorded in flat file format or database.

Though, it is apparent that there are many problems from manual data handling, the major problem is that data and information is not updated in a structured manner when it needs to be. In responding to a question about monitoring vital signs data updating, respondents mentioned that , there is no proper mechanism to update this data item. This is other indication of the possible underreporting scenarios.

A question about the level and type of patient monitoring analysis was another important issue raised. Given the inefficacy of data handling system and unavailability of complete quality data, rigorous analysis and reporting is impossible. Basically, the analysis task is focused on descriptive analysis, which also done by the doctor without recording the results. To respond to the "how many" question. All respondents are agree and expressed that; "There is no statistical analysis to the patient vital signs measurement sheet form".

This is also supported by the observation made by the researcher on the selected research sites and the measurement sheet form being used at the healthcare departments. It is easy to learn that very high level, summarized analysis of patient monitoring data prohibits from making maximum use of the knowledge hidden. It is also learned that severity analysis, identification of chronic patient status factors, and classification of patients disease analysis were mentioned as analysis requirement to decision support in healthcare monitoring domain. These entails the requirement for multidimensional new way of patients monitoring data analysis that provides maximum use of the knowledge accumulated.

5.8.4 People and Motivation in healthcare monitoring Information Management

Questions regarding people and motivations in healthcare monitoring data collection and analysis include: What is the prime motivation of managing patient vital signs monitoring data? . What sort of information is required by stakeholders regarding healthcare monitoring? Who do you think is responsible in patient vital signs monitoring data reporting? Who do you think should participate in patient vital signs monitoring data management and analysis? Who are the primary users of patient vital signs monitoring information?

In response to the above questions, interviewees clearly stated about the need to develop evidence based hospital healthcare monitoring plan is the prim motive for the management of patient vital signs monitoring data. A respondent from healthcare monitoring department stressed that this will happen by coordinating with others departments. Another respondent forwarded another important strategic motivation to promote participation of the relatives in healthcare monitoring. But, it is noted from the discussion the fact that the data is incomplete, the absent of data analysis and reports prohibited the effort to achieve the healthcare monitoring objectives.

It has been identified that there are various stakeholders, who directly or indirectly need the patient's vital signs monitoring information. However, there is no formal exchange of detailed of patient's vital signs monitoring data between healthcare department, doctors, and others departments to produce general real time statistics analysis. The researcher observed that most of the respondents are in favor of this idea.

5.8.5 Time and Network of healthcare monitoring Information Management

Location and network along with the timing issue is another theme identified. To investigate the issue the following questions are forwarded: Can you mention a specific place (like organizational units) where data reporting, analysis and dissemination should happen? Would you comment please on the timing for health monitoring data reporting, analysis and dissemination?

A number of respondents mentioned that these issues are critical and can significantly affect the learning process. In responding to the first question, there was considerable confusion over who should take on the primary responsibility of patient vital signs monitoring data handling. As has been in many countries healthcare monitoring departments in hospitals are currently responsible of patient vital signs monitoring data collection, but with absent of data analysis. This claim sounds logical, as healthcare monitoring departments are mainly responsible in reporting patient vital signs monitoring data, which will be difficult without proper and multitype analysis. In connection to this, the research at hand argues that though there might be other administrative or architectural solutions.

Designing an **AmI** healthcare decision support and monitoring architecture could solve this problem. The architecture will allow keeping real time monitoring and real time analysis and classification of patient's status and alert alarm when the situation of patient is chronic. The type of analysis at this site is expected to be a detail one and to make use of the classification, exploratory machine learning and data mining capability.

Regarding the timing, it was found that patient vital signs monitoring data should be reported as soon as possible so that relevant information will affect the learning will not disappear and appropriate health surveillance happens. Real time update of patient vital signs monitoring data records was one of the recommendations by respondents. It was also noted that real time analysis with on-demand exploratory classification analysis are fundamental and the dissemination will follow through various channels.

5.9. Summary of the Chapter

This chapter presented the foundations of the AmI healthcare monitoring information architecture defined in this research. It started with a general overview, which is followed by the results of experimental machine learning approaches in addressing a portion of the objectives set for the whole research. Drawing from previous data mining researches in analyzing healthcare monitoring situations, the current research is focused on determining role of healthcare monitoring department and related factors on patient vital signs monitoring data. This is then used to populate the content dimension of the proposed information architecture exploration of data quality issues and patient monitoring analysis trends also inform the function dimension of the architecture. Moreover experimenting on ensemble models and comparisons are also part of the task. A qualitative analysis of healthcare patient's vital signs information management practice in Khartoum State hospitals was also presented in the third sub-section of this Chapter. Its focus was on the problems and information requirements of the patient vital signs monitoring information management. By presenting this result, in addition to providing a ground for the definition of an information architecture presented in the next chapter, a couple of objectives set for this research were also addressed.

6 Ambient Intelligence Healthcare Monitoring Information Architecture (AmIHCMIA)

6.1 Overview

From the literature review and combining with the results gained from the analysis of a qualitative empirical data and machine learning experiments as reported in the previous Chapters, this Chapter proposes integrated information architecture for AmIHCM domain.

More specifically, the characteristics identified through Data mining experiments, data quality problems and the results of qualitative data analysis are used in developing architectural description under the six dimensions.

Accordingly, one of the research sub questions of this Thesis, mentioned in chapter one and as mentioned below, is addressed using theories and empirical results from previous Chapters.

- *How generic Information architecture can help in developing the **AmI** assisted health care architecture?*

  In addressing this research question, we define two sub objectives as mentioned in chapter one and presented below to guide the process.:
- *To investigate a way to define information architecture based on enterprise architecture framework (EAF) to establish **AmI** healthcare monitoring decision support and monitoring information management.*
- *To Define and develop a **AmI** healthcare monitoring decision support and monitoring information architecture that will facilitate effective utilization and management of healthcare monitoring information among hospital departments.*

Rating of the existing healthcare monitoring scenario and practices in healthcare monitoring information management (IM) and a literature review and technologies in the area provide a base to the formulation of information architecture guided by known enterprise architecture framework (EAF). Moreover, the empirical result of knowledge discovery experiments conducted on the simulated wearable sensors dataset in Elbaraha medical city environment, also disclosed some important issues including analysis requirement and data quality that needs to be addressed. The experiments also identified effective classification models that integrated to the healthcare vital signs monitoring data system as per the architecture. Accordingly, a description of the defined general architecture followed by the proposed improvements is presented below.

6.2 The AmI healthcare monitoring Information Architecture

According to literature review and existing practices in a local context and result of our successive data mining experiments , illustrated the need for an integrated approach to the **AmIHCM** data collection and analysis, which will be addressed by an enterprise view. Therefore, in order to define an information architecture model, two dimensions of people and operative perspectives are mentioned based on the **ZF** .

The first dimension, the viewpoint of the people, are the actors of **AmI** healthcare monitoring data collection and analysis process. The second one, operative perspectives, defines the several requirements, constraints and operations that should be created to the architecture of the AmI healthcare monitoring data collection and analysis. In the process of the proposed architecture, the essence of these dimensions is described. The content of each cell that is deduced from the crossover of columns and rows is also defined, as described in . Since **ZF** is more general, including different views and dimensions at different levels, it is relevant to delimit its use as per the research scope under consideration. It is essential to get a clear definition of the contents of every cell. The **ZF** is simply a framework. It is not a process, a method, a notation or a tool .

Accordingly, computation (Rows 1 and 2) and platform independent (row 3) representations are selected because of the scope of the research and to keep the flexibility of the implementation of the architecture. Similarly, based on their relevance to the domain and scope of the research, the six dimensions or columns of the **ZF** are emphasized by . In , four dimensions were suggested namely, the what, why, who and how as the minimum set of descriptions for each representation, though, it may be necessary to add more descriptions depending on the business domain under study.

Thus, reconsidering the idea given in , for our case **AmIHCM** data collection and analysis, information management could be represented in a 3 by 6 matrix containing the top three viewpoints (rows) and six dimensions (columns). The rows of this architectural model describe different viewpoints to approach the subject at different layers; and these rows are described in Table 6.1.

Table 6.1: Rows of the AmIHCMIA model

View Points	Description
SCOPE (CONTEXT)	Describes the strategies, content and constraints of the AmI health care monitoring, data collection and analysis.
BUSINESS (CONCEPT)	Define goals, structure and processes that are used to support mission of AmI healthcare monitoring organizations with respect to healthcare monitoring information as an enterprise.
SYSTEM (LOGICAL)	Describes system requirements, objects, activities and functions, network that implements the business model. The system model states how the system is to perform its functions.

Similarly, the columns of this architectural model describe several dimensions of the model, which are referred to as abstractions; and these columns are explained in Table 6.2.

Table 6.2: Dimensions of the AmIHCMIA model

Dimensions	Descriptions
WHY (MOTIVATION)	It translates AmI health care monitoring strategies and objectives into specific meaning.
WHO (ENTITIES)	It defines who is related to AmI health care monitoring data and information management.
WHAT (DATA)	It is data to define and understand AmI healthcare monitoring
HOW (PROCESSES)	Processes to translate AmI healthcare monitoring requirements into more detailed implementation and operation definitions.
WHERE (PLACE)	It is related to physical distribution of places where AmbI healthcare monitoring data collection and analysis will be implemented and operated.
WHEN (TIME)	It describes how time influences AmI assisted health care monitoring information management.

According to the above discussions, a 3 x 6 grid representation of the architecture is presented in Tables 6.3 and 6.4, by identifying relevant dimensions and user views for **AmI** healthcare monitoring data collection and analysis practice. Accordingly, a higher level AmI healthcare monitoring Information Architecture is offered based on the **ZF**. As indicated on the architecture (Tables 6.1 and 6.2), contextual or the first row, defines higher-level view of the data, people, function, time, place and motivation of **AmI** healthcare monitoring data collection and analysis. This creates an umbrella under which the conceptual and logical design of the **AmI** healthcare monitoring data collection and analysis is to be implemented.

Returning to the columns, the "Why" enables to clearly state the motivations at the three levels (rows), the 'what" will describe, what to collect, analyze and disseminate, the "how" help in defining how to collect, analysis and disseminate, the "When" will guide the timing, the "where" depict the geographical distribution of activities and lastly the "Who" depict who should collect, analyze, use and disseminate. The next level of architecture produces various drawings, checklists, diagrams appropriate in representing the above-mentioned content with in their context for their respective users.

Table 6.3: AmIHCMIA based on Zachman approach - I

	Why (Motivation)	**Who (People)**	**What (Content)**
Scope (Contextual)	Service care delivery in healthcare monitoring and impact to personal and public health	Essential Healthcare monitoring service organizations and their functions.	Description of important healthcare monitoring service and care delivery information.
Enterprise and Environment(Conceptual)	Personal healthcare monitoring benefit and care monitoring delivery business objectives.	Healthcare monitoring information system workflow	Semantic description of health care monitoring processes
AmI healthcare monitoring System (Logical Design)	System functional requirements.	Healthcare monitoring information system human system interface architecture.	Logical data model for healthcare monitoring information.

Table 6.4: AmIHCMIA based on Zachman approach - II

	Where (Network)	**How (Function)**	**When (Time)**
Scope (Contextual)	Identification And description of organization and individual locations.	Important healthcare monitoring and monitoring care delivery services.	Identification of significant health care monitoring and care monitoring delivery events.
Enterprise and Environment (Conceptual)	Structure and interrelationship of health care monitoring facilities.	Conceptual activity model of health care monitoring delivery.	Sequence and timelines of healthcare monitoring services.
AmI healthcare monitoring System (Logical Design)	Connectivity and distributed system architecture	Application architecture with function and user views.	Healthcare monitoring event phases and process components

Given the high level architectural representation, the detail description is given below: Accordingly, based on the empirical results, literature review practice and a method to define an enterprise architecture using **ZF** is proposed in . A description of an architectural model for AmI healthcare monitoring data collection and analysis is presented by views/perspectives.

Contextual View- Scope/Planners Perspective

This perspective can be seen as a means of defining the scope of the expected system. It is concerned with the issues of the planning of the healthcare monitoring data collection and analysis system. There is no dependency among cells' concepts according to Pereira and Sousa, . So, the order of fulfilling for this row is totally free. Drawing from interview results presented in Chapter 5 and analysis of a **AmIHCM** strategic plan was also used to define the content of cells at this viewpoint.

Accordingly, the first cell (Context-Motivation) addresses public and individual health monitoring and the business of healthcare monitoring delivery across enterprise boundaries. The content is presented in Table 6.5. Similarly the next cell (Context-People) Identify the essential components of the health care delivery system. Essential healthcare monitoring entities or organizations identified, while the third cell (Context-Content) Identify and describe the important global healthcare monitoring service and healthcare monitoring delivery information. Data hierarchy can be extracted based on data analysis of enterprise strategic plan, .

The fourth cell (Context-Network) Identify and describe the global entities involved in delivering healthcare monitoring services. Individual and organizational participants in the organization. Addresses the location of an enterprise . The fifth cell (Context-Functions). identify, describe, and regulate important business operates and delivery services. The sixth cell (Context-time) Identify and describe the fundamental organization delivery events independent of profession, specialty or services delivery environment .

As indicated above and recommended in the empirical (Section 5.3), data mining results (Section 5.2), the researcher's own observation are analyzed in the process of populating the first row of the model as presented in Tables 6.7 and 6.8. Motivations, participating entities and accident data content are presented in Table 6.7 while the place, process and timing aspect of accident data is exhibited in Table 6.8.

Table 6.5: Contextual view- scope/planners perspective

	Motivation	**People**	**Content**
Scope (Contextual) AmI healthcare monitoring	▪ Develop Efficient AmI assisted health care monitoring, decision support and monitoring in a form of goals ▪ Reduce the cost for the tradicional healthcare monitoring. ▪ Provide efficient medical services and improve the health care monitoring.	▪ Essential AmI assisted health care monitoring organizations ▪ AmI healthcare Manager Coordination Office ▪ Patients with wearable sensors ▪ Doctors Office ▪ Nurses Office ▪ Relatives	▪ Monitoring Patients (with wearable sensors). Medical services. ▪ Doctors with PAD treats patients, receives Alarm and messages ▪ Nurse with PAD Medical services. receives Alarm and messages ▪ Manager ▪ Manages the DB With computer work station, decision support, send Alarm and messages ▪ Environment
	Why	Who	What

Table 6.6: Contextual view- scope/planners perspective

	Network	**Function**	**Time**

Scope (Contextual) AmI healthcare monitoring	▪ Locations of patients data collection and analysis system ▪ Patients wearable sensors (Rooms) ▪ Nurses offices ▪ Doctors Office ▪ Manager office ▪ Relatives.	▪ Patients sensors data collection and reporting ▪ FRID and Access Points ▪ Quality and completeness checking, ▪ Updating monitoring data ▪ Sending Alarm and Messages ▪ Machine learning and data mining analysis (Descriptive analysis, Exploratory analysis. ▪ Novel Ensemble Decision Support and Health Care Monitoring System ▪ Decision support assisted	▪ Monitoring events ▪ Monitoring recording ▪ Data Quality check ▪ Monitoring analysis ▪ Alarm and messages ▪ Decision support. ▪ Dissemination
	Where	How	When

Conceptual View-Enterprise/Environment Perspective

The second important users view considered was conceptual view, which is from the enterprise and environment aspect. Based on the recommendation , cell contents at this level are derived from the above contextual perspectives. This perspective models the motivation, people, content, network, time and functions of an **AmIHCM** data collection and analysis system from the viewpoint of business owners.

Accordingly, the first cell (conceptual-motivation) identifies and describes the means to quantify individual healthfulness and the business objectives of a healthcare monitoring delivery organization. The major reason for this specific system from enterprise point of view is presented in terms of objectives. The second cell (Conceptual-People) Identify and define the roles of individuals participating in healthcare monitoring delivery in an organization. Table 6.7 allows exhibiting the general work flow in between and within **AmIHCM** organizations. As explained in , Organization chart or Processes vs. Organization Matrix can be used to model the content of this cell.

Table 6.7: Conceptual-people using process vs. organization matrix

	HCMDC	QCU	DA	PA	Al	ISD	TR	MS

Manager	X	X	X	X	X	X		
Patients with wearable sensors	X					X		
Doctor							X	X
Nurse							X	X
Relatives					X	X		

Key: *HCMDC: Healthcare monitoring data collection (reporting and referencing), QCU: Quality Checking and Update, DA: Descriptive analysis (interpretation), PA: Predictive & Exploratory analysis (interpretation), AL: Alarm, TR: Treat patients, MS: Medical Service ISD: Information Sharing and dissemination*

The third Cell (Conceptual-Content) presents and describes the essential types of information required for operation of a care delivery organization, which also includes semantic description of the AmI healthcare monitoring data. The business contents include business entities, their attributes and relationship. Entity dictionary can be used to represent this cell .The fourth cell (Conceptual-Network). specifies and describes the layout of health care monitoring facilities and their interconnection. The location of business nodes, where the system is used by Ertaual and Sudarsanam . It is also mentioned in that, the focus of this cell is to represent the conceptual model of “Where”, which includes the location of and place, where stakeholders, use from the system and also the operations that they can do related to this.

As described in , organizational units within location stereotypes of **UML** packages associated using dependency relationship is a preferred modeling to represent this cell. The content of this cell is represented accordingly as shown in Figure 6.1. As evident from Figure 6.1, five important AmI healthcare monitoring information management locations are represented with their dependency relationship. It also agrees with the proposed approach , that considers the content of conceptual-function and contextual-network cells to populate this cell.

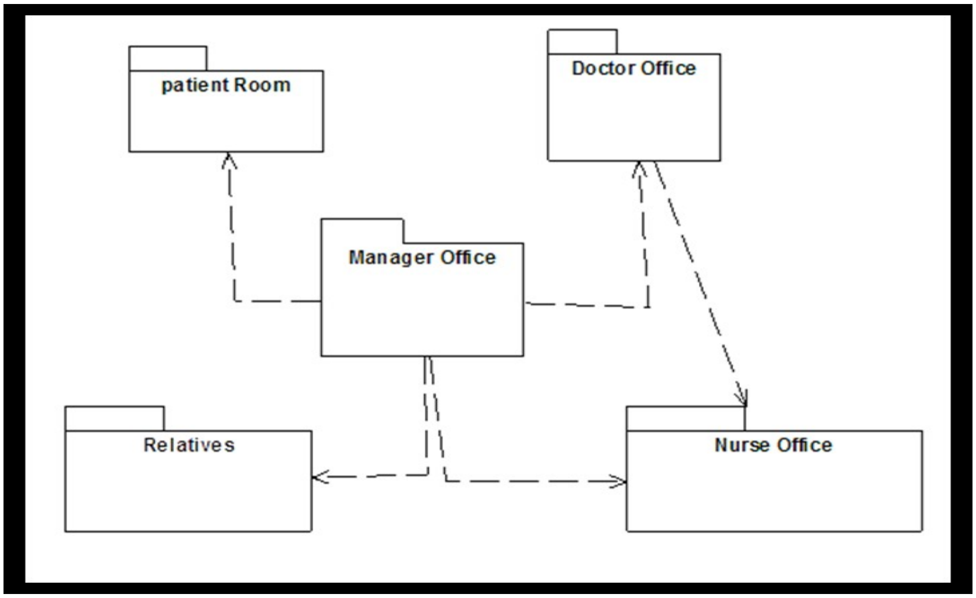


Figure 6.1: UML package representing conceptual-network cell

The fifth cell within this perspective, (Conceptual- Function) identifies and describes the fundamental health care monitoring, management and support activities in a care monitoring delivery organization. Models the business workflow of the stakeholders interacting with the business. Flowchart, activity diagram, **UML** use case diagram and sequence diagrams are common tools for process modeling at this layer .

Accordingly, activity diagram is used to represent the content of this cell as depicts in Figure 6.2. It clearly explicates the workflow between the locations identified earlier. Its content is derived from the cell in the above row, contextual-function.

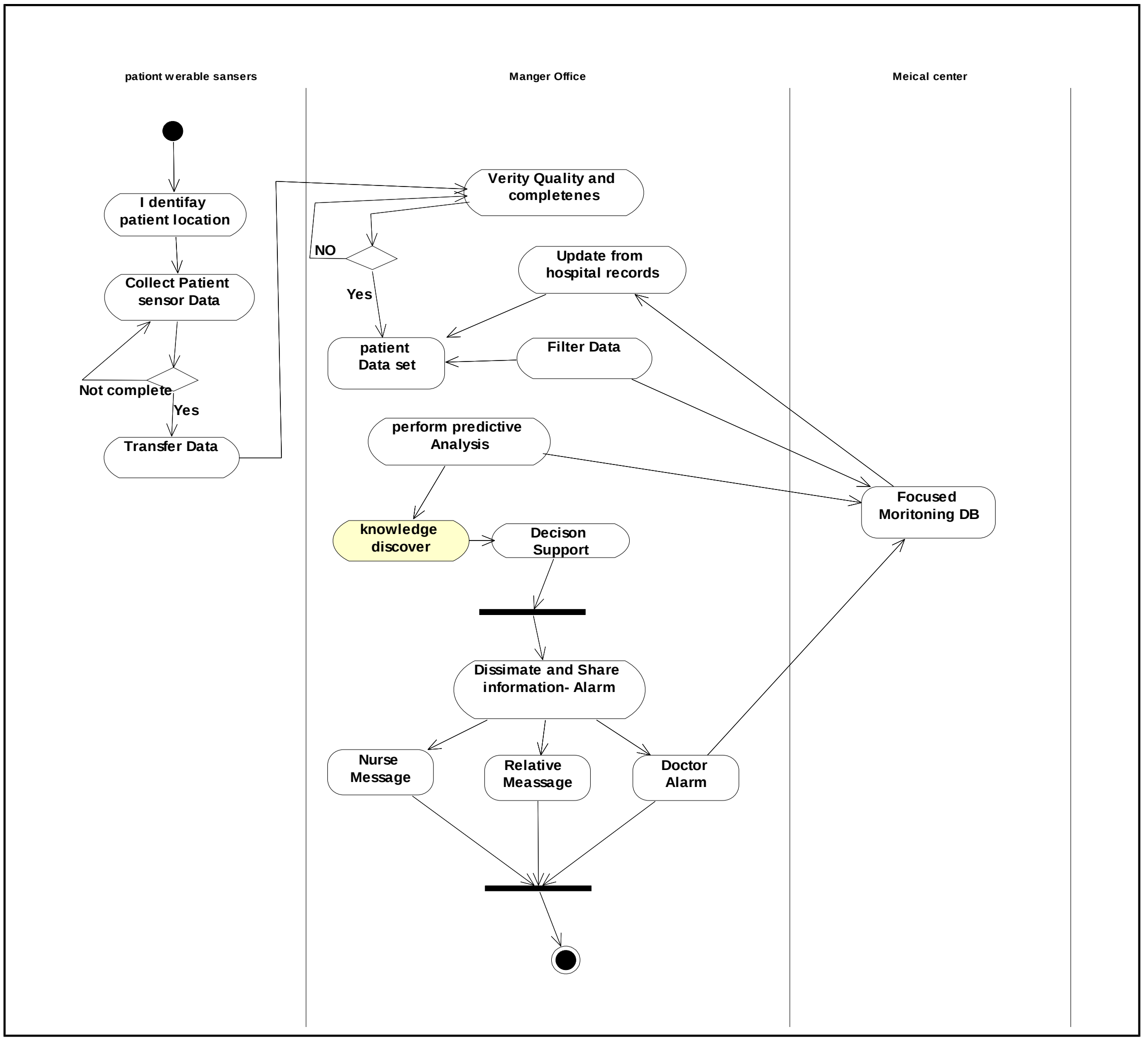


Figure 6.2: Activity diagram representing conceptual-function cell

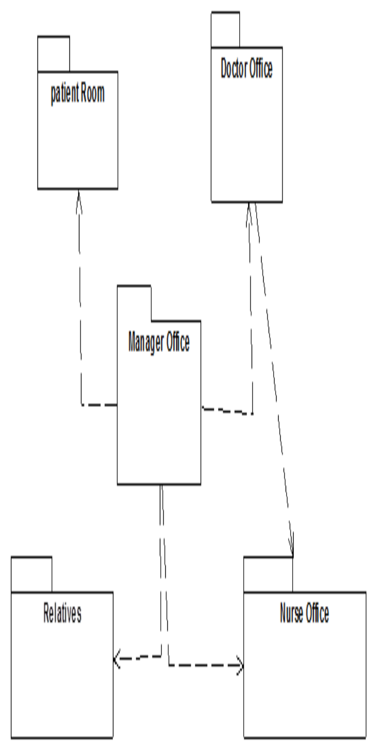
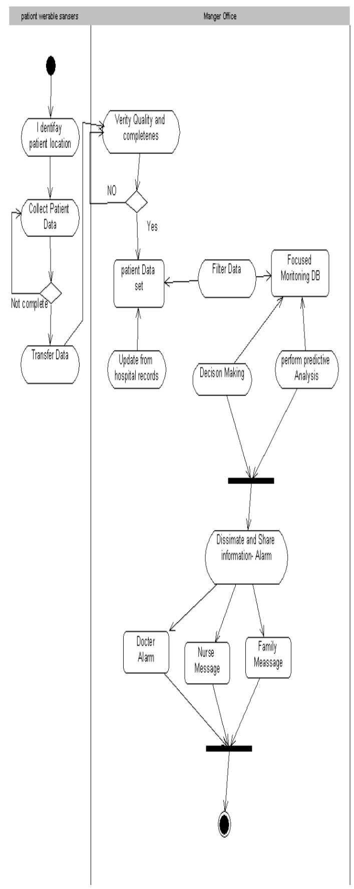
Finally, the sixth cell (Conceptual-Time) focuses on determining the order and timing for the processes of fundamental healthcare monitoring services in a care delivery organization; sequencing of the timing of process, events and flows significant to the business, the healthcare monitoring data collection and analysis. According to Ertaul and Sudarsanam, , time dimension may be of two forms. One of the forms represents the snapshot of a point in time and the other defines a period.

There is a number of representations like Business execution plan, GANTT, PERT charts and sequence diagram using business actors, were proposed in the literature . Hence, a list of events with their suggested period is a preferred approach for its understandability to represent this cell. Thus the populated architectural description of the second row of the **AmIHCM** information management system is presented in Tables 6.8 and 6.9. The conceptual level representations of motivation, people and content are shown in Table 6.8 while Table 6.9 contains models and description for the network, function and time dimensions.

Table 6.8: Conceptual view-enterprise/environment perspective

	Motivation	**People**	**Content**
Enterprise and Environment (Conceptual)	▪ Single Complete dataset Source of Information. ▪ Distributed system with dynamic , multi-view ▪ Advanced Analysis data mining and machine learning methods ▪ Decision support system. ▪ Facilitated information Sharing	1.Manager Office: ▪ Monitoring data collection (reporting and referencing). ▪ Data mining and machine learning analysis methods. ▪ Decision support ▪ Information Sharing and dissemination ▪ Sending Alarm and messages 2. Nurses office. ▪ medical services. ▪ Receive message. 3. Doctors Office ▪ Treats patients ▪ Receives Alarm and messages	▪ Monitoring vital sign ▪ Patient sensors; personal details ▪ Room no ▪ Sensor id. ▪ Sensors Vital signs; types, nature. ▪ Environment; weather condition, temperature. ▪ Light condition, time ▪ Doctor: ▪ Specialist in …. ▪ Receives Alert ▪ Phone no ▪ Nurse ▪ Location ▪ Phone no. ▪ Receives alert and messages
	Why	Who	What

Table 6.9: Conceptual view-enterprise/environment perspective

	Network	Function	Time
Enterprise and Environment (Conceptual)	 Figure 1 : UML package representing conceptual-network cell	 Figure 2: Activity Diagram representing conceptual-function cell	▪ Monitoring events ▪ Critical patients situations. ▪ Decision support ▪ Monitoring recording (immediately) ▪ Vital signs monitoring Analysis ▪ Descriptive analysis (on the spot). ▪ Data mining and machine learning
	Where	How	When

Logical View- AmI healthcare monitoring Data Collection and Analysis System Perspective

This perspective models the requirements, participation, business content, and process of the **AmIHCM** data collection and analysis system from the viewpoint of a system. It helps to give specific functional requirements view of the system, human-system interface issue, logical data model, geographical location, timing and layered architectural design of the data collection and analysis system with functions and users views. Architectural artifacts of this perspective are presented in Tables 6.10 and 6.11. Accordingly, the first cell (Logical-Motivation) relate to the functional requirements and the test and acceptance criteria for a healthcare monitoring information system.

This also presents the reason of the system in terms of functional requirements. Using data from the interviews and review of existing healthcare monitoring initiatives, it was possible to develop functional requirement of the **AmIHCM** data collection and analysis system expressed as behavioral objectives to populate the Motivation (Why) component of this perspective.

In , it is recommended to consider the analysis of the cells above and the content of logical-function cell in defining the content of this cell i.e Logical-Motivation. The second cell (Logical-People) Detail the methods used for the description of the functioning architecture for the interaction of individuals with the healthcare monitoring information system. This also describes the structure and contents of user interactions with the system. It can be modeled using Systems vs. Roles Matrix or **UML** Use Case diagram .

Accordingly, Use Case model as depicts in Figure 3 is used to represent the proposed user interaction in a **AmI** healthcare monitoring information management scenario. The inclusion of relatives in the process of **AmI** healthcare monitoring information management is worth mentioning. Its content is derived from the above cell and analysis of the function perspective of the systems view.

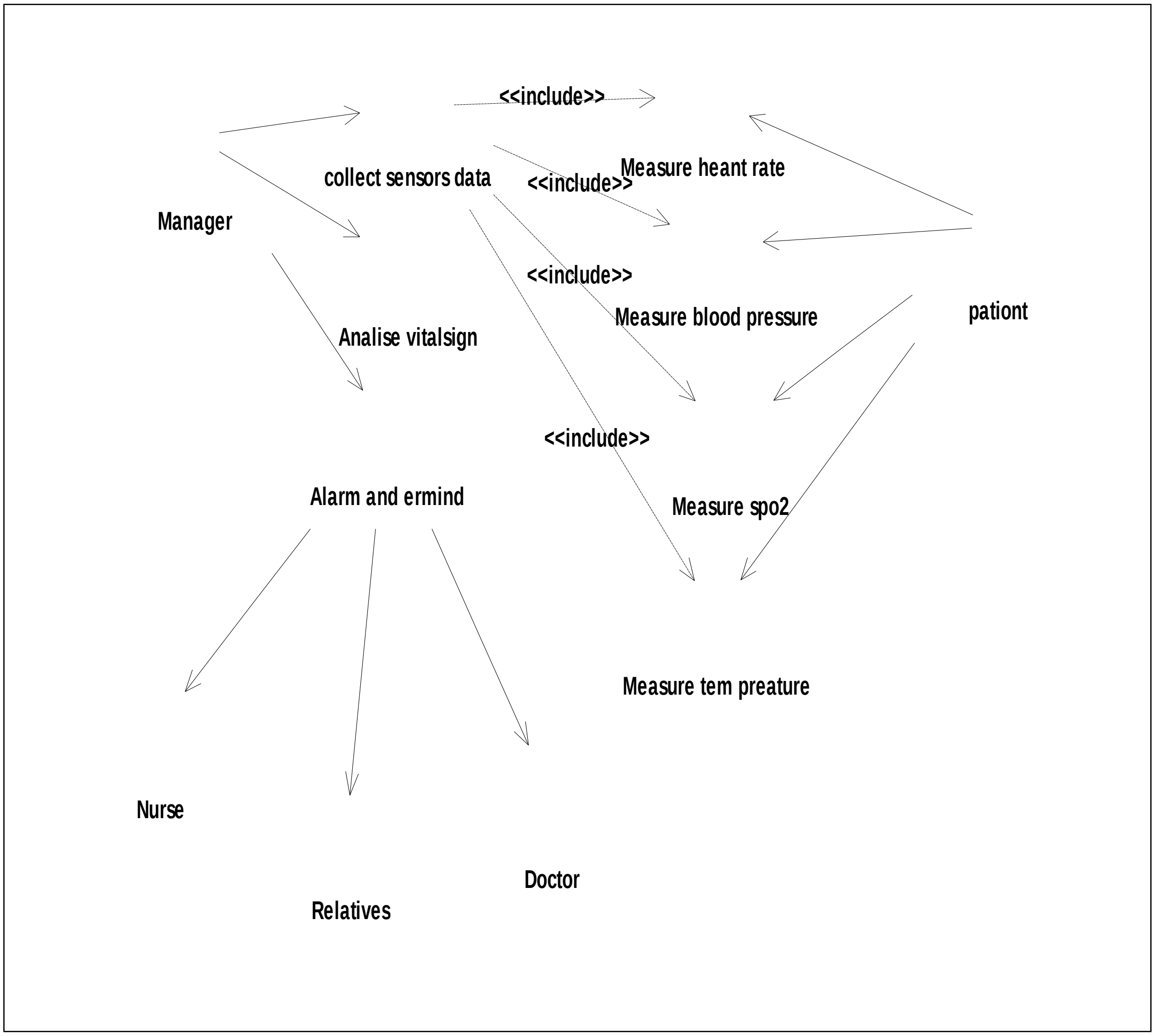


Figure 6.3: USE CASE Model representing logical-people cell

The third cell (Logical-Content) Detail the methods used to prepare a logical data model, or the non-technological description of the data used for healthcare monitoring delivery in an organization.

This describes the business structure including business entities and their relationship. Example model of this cell could be a business entity-relationship diagram (ERD) that models the business concepts, entities and attributes. The content of this cell could be extracted from owner's perspective output . It is depicted in Figure 6.4.

Figure 6.4: E-R model representing logical-content cell

The forth cell in this perspective (Logical-Network) Describes the communication architecture supporting healthcare monitoring and care monitoring delivery. This represents the available nodes of a whole system and logical links in between them. In the healthcare monitoring domain the Patients Sensors, Relatives, Doctors office, nurse office and Manager Office are modeled with their respective modules are modeled.

This cell has the conceptual-network and logical-function artifacts as its base . Though, system diagram and **UML** component are also proposed, Deployment diagram, as shown in

Figure 6.5, using location stereotype of packages is a preferred modeling techniques to represent this cell .

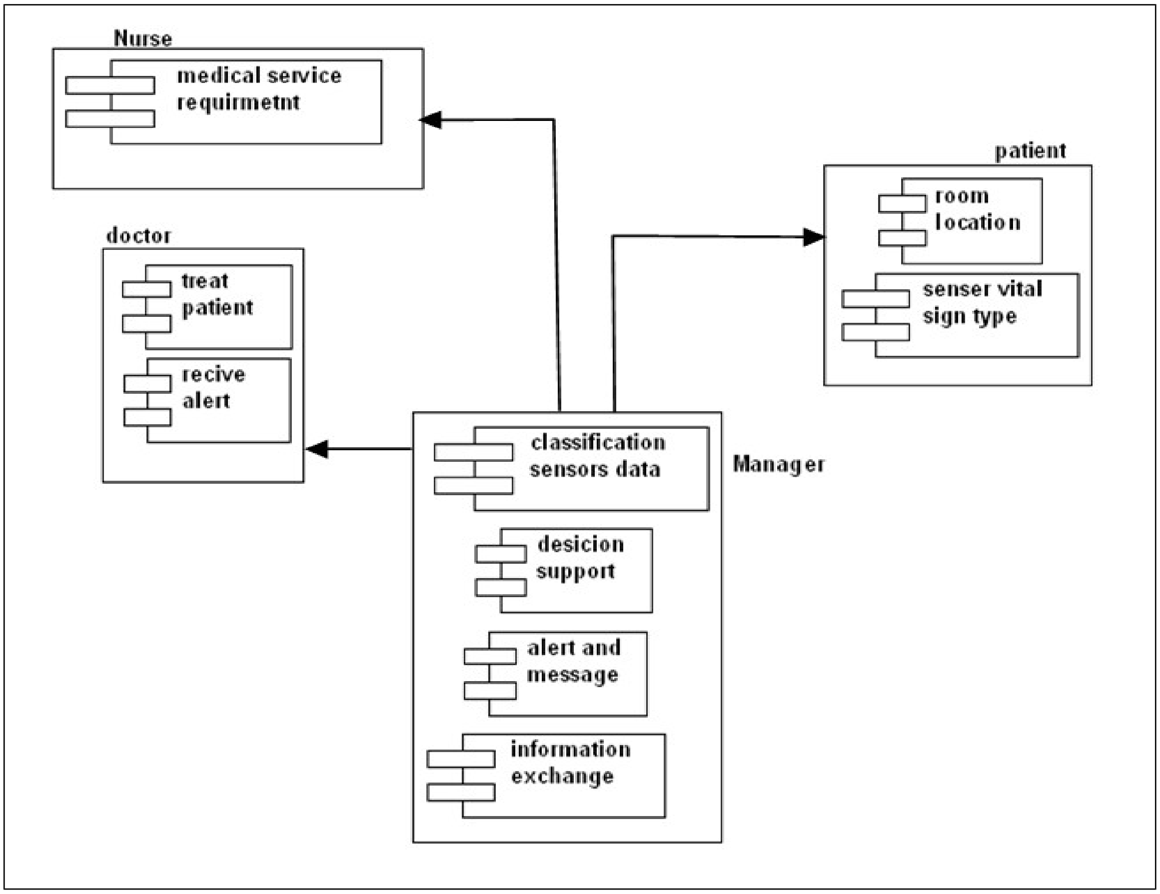


Figure 6.5: Deployment diagram representing logical –network

The fifth cell (Logical-function). Describe the structure of software to support healthcare monitoring and care monitoring delivery processes. This represents a layered architectural design of **AmIHCM** data collection and analysis system with function and user views . Its content can be derived from the cell immediately above. Thus, it specifies the structure, the responsibilities and the relationships of the design elements of a healthcare monitoring data collection and analysis system. As can be seen from the diagram Figure 6.6 depicts the system in three layers, where the bottom layer represents the data storage while the upper two layers illustrate the business logic and the presentation or end users view respectively.

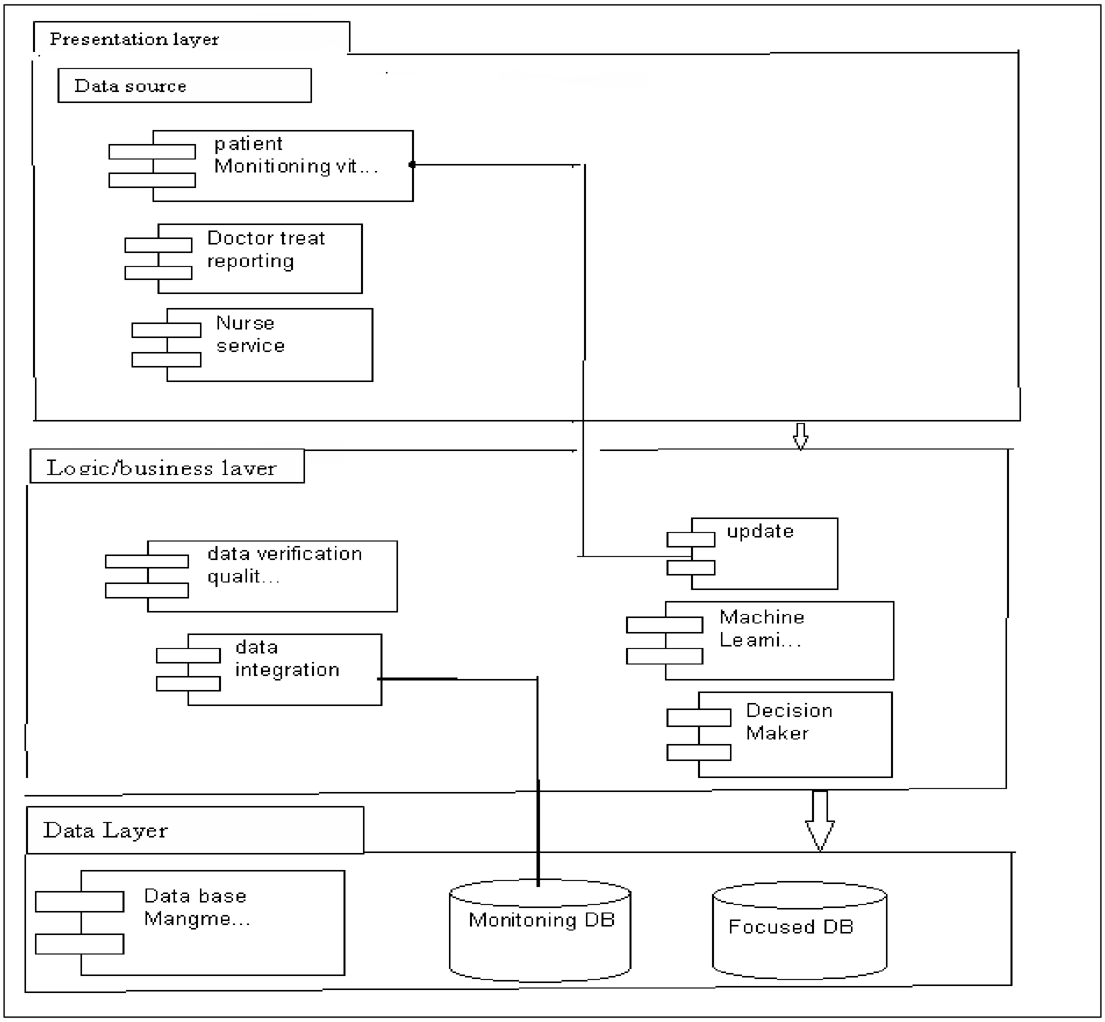


Figure 6.6: Layer architecture representing logical – function cell

Finally, the sixth cell (Logical-Time) Detail the methods used to describe or descriptions of processes and event sequences within a care delivery organization. This is used to represent events with their sequence and period in a more detail. State diagrams and sequence diagrams are recommended in the literature to define the content of this cell . Sequence diagram depicted in Figure 6.7 uses **UML** notes for periods is a suggested diagram to represent timing in **AmIHCM** information management . The diagram explicates the process of **AmIHCM** data collection and analysis practices emphasizing on time sequencing of the events.

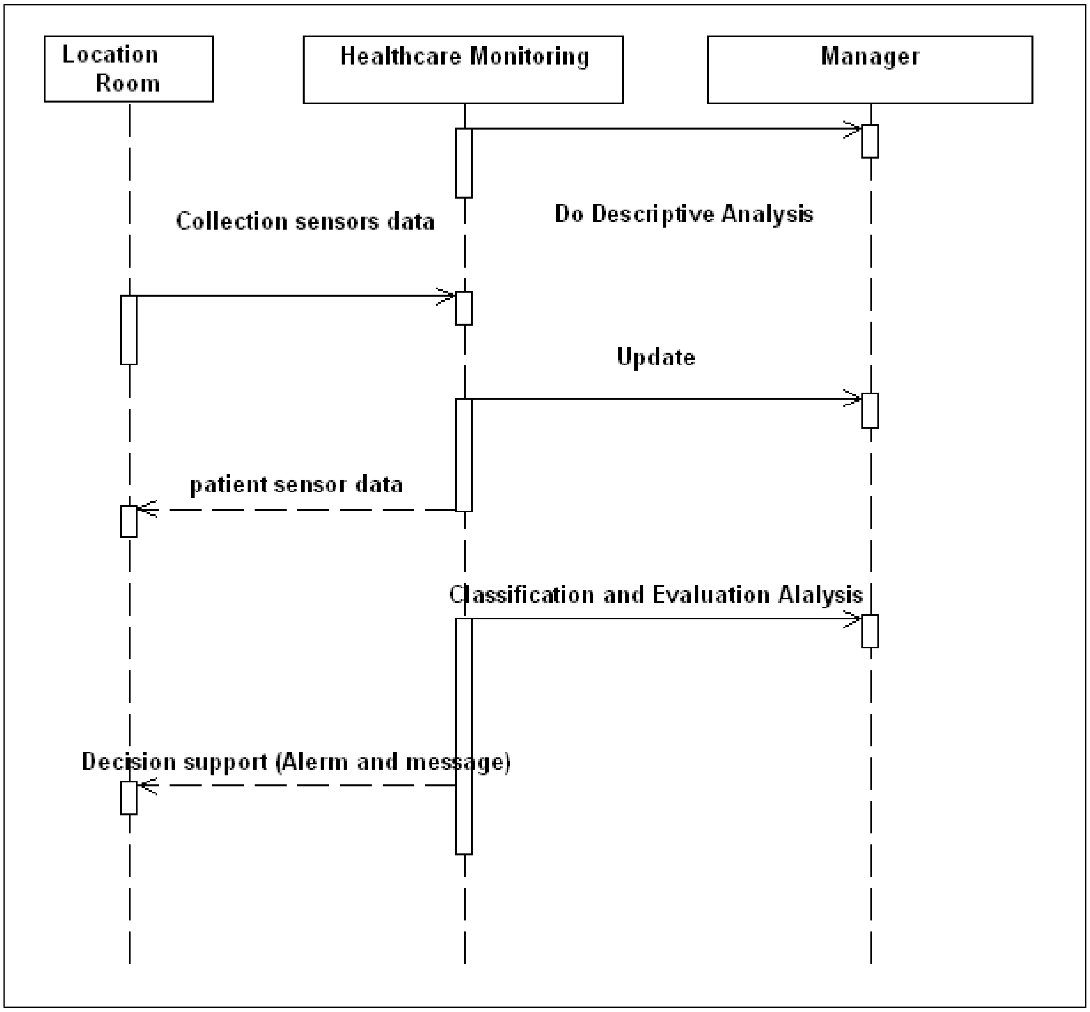


Figure 6.7: Sequence diagram representing logical– function cell

The logical view of the **AmIHCM** data system is shown in an integrated manner in Tables 6.12 and 6.13. The integration is expressed in terms of coherence in between the contents of each cell. Accordingly, the logical level representations of motivation, people and content are shown in Table 6.12 while Table 6.13 contains models and description for the network, function and time dimensions

Table 6.10: Logical view- HCM data collection and analysis perspective

	Motivation	People	Content

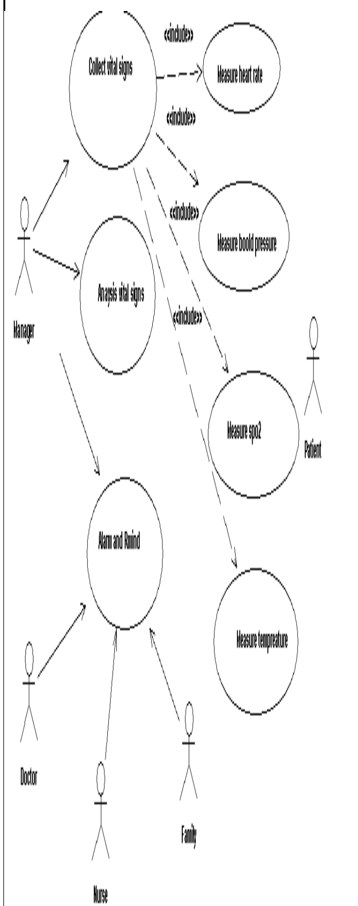
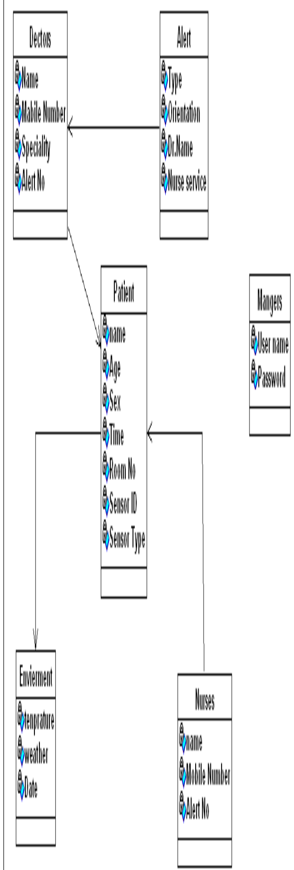
AmI healthcare monitoring Data Collection and Analysis System (Logical Design)	▪ Functional requirements of AmI data collection and monitoring analysis system ▪ Quality and accurate information ▪ Provide AmI assisted healthcare and monitoring platform for decision support in real-time. ▪ Information sharing ▪ Enable periodic data mining and machine learing analysis and techniques ▪ Enable exploratory and Decision support.	 Figure 6.3: Use case Model representing Logical-People cell	 Figure 6.4: ER Model representing logical-Content cell
	Why	Who	What

Table 6.11: logical view- HCM data collection and analysis perspective

	Network	Function	Time

<table>
<tr>
<td>AmI Healthcare Data

Collection and

Analysis

System (Logical

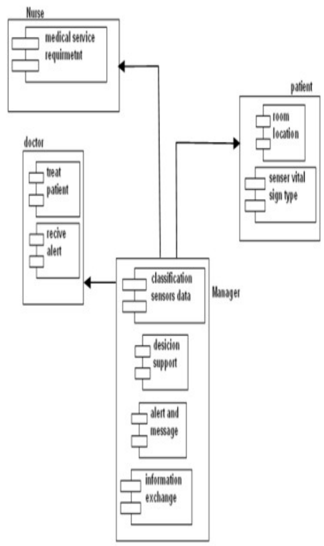
Design)</td>
<td>

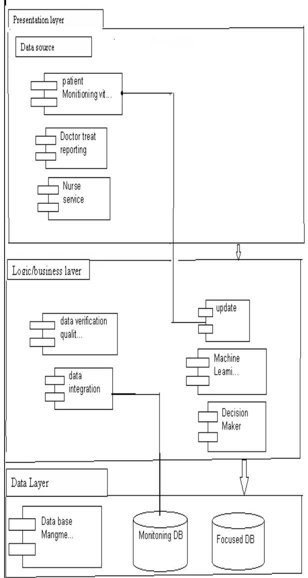
Figure 6.5 : Deployment diagram representing logical –network</td>
<td>

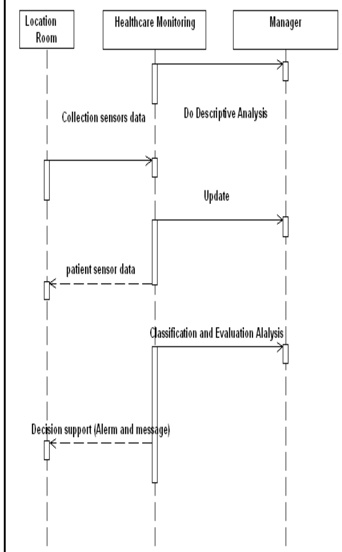
Figure 6.6: Layer Architecture representing Logical – Function cell</td>
<td>

Figure 6.7: Sequence Diagram Representing Logical– Function cell.</td>
</tr>
<tr>
<td></td>
<td>Where</td>
<td>How</td>
<td>When</td>
</tr>
</table>

6.3 Discussions

In this Section, we discuss the main features of the proposed AmI healthcare monitoring IA. The architecture presents a mechanism to facilitate communication and integration regarding healthcare monitoring data by which it satisfies one of the major purposes of architecture. Based on the generic **ZF**, the **AmIHCMIA** is represented by the integration of rows and columns.

The framework decomposes enterprise architectures into level of cells by column and row that reduces complexity. Column represents a specific dimension each column at each row point has a unique model, while each row at each column point presents a unique perspective. Representation of the architectural descriptions is based on the data and process dimensions, identified in the data mining experiments and a qualitative data collection and analysis process.

We can note that each perspective could be applied to both development of new systems and maintenance of existing systems. Effective enterprise information architecture is more than just populating Zachman cells, is that architecture which integrated data, technology and application. In line with the argument directed in , without an integrated perspective, all architectural models and strategies will be, inflexible and will remain unused.

Thus, in this research by applying the top three levels of the Zachman hierarchy, it was possible to develop **AmIHCMIA**. Major procedures includes; determining the dimensions and perspectives required, determining the information requirements of the domain through qualitative data collection and data mining experiments, classifying and structuring the information based on the dimensions identified keeping the integrity, representing the information in a required level of detail for the perspective under consideration, evaluating both the process and the content using recommended metrics and methods. Therefore, though one of the objective of this research is to design the AmI healthcare monitoring information architecture of healthcare monitoring organizations based on a **ZF**, the process and techniques employed as a methodology is another major contribution of this work.

In analyzing the framework's perspectives, the authors , verified the existence of a concept related to the ZF designate as an "anchor cell". According to the authors, an anchor cell is a cell that on any framework's perspective has an aggregate function relative to the others cells. In line with this, it is easy to learn that two cells representing the intersection of the conceptual and logical perspectives under the function or process dimension are more of aggregate.

In the context of this specific research, the description of content, process and people at a conceptual and logical level satisfy the 'anchor cell' concept. This is mainly because these cells represent the major issues in a healthcare monitoring information management practice under study. The concept of 'anchor cell' agrees also with the . Information Architecture core representations; content, context and users.

A key distinguishing feature of **AmIHCMIA** is simplicity. Partly, the simplicity come from the nature of the guiding framework selected. Architectural model is a user practical integrated architectural model that balances ease of adoption and use with completeness and theoretical stringency.

The Objective of the **AmIHCMIA** is to assist in evaluating or developing healthcare monitoring data collection and analysis systems and promoting informed decision support by healthcare monitoring organizations. **AmIHCMIA** attempts to concentrate on the best procedures, practices and artefacts that have inherent in several countries, as learned from literature supported by empirical investigations in the Khartoum state case and generalize these into a general architecture that can be applied in whole Sudan and other similar environment countries.

6.4 Conclusions

AmI healthcare monitoring Information Architecture, which is the major deliverable of this research work, has been presented in detail in this Chapter. The discussion includes both the process and the end results. It starts with an overview explaining the link that this chapter has with the previous Sections. Then the architecture is presented and guided by an enterprise architectural framework. The discussion followed also provided details of its applicability and key features of the architecture.

Thus, the development of the **AmIHCMIA** in an enterprise view is an original contribution, which improves and expands the conceptual framework of the research in the domain. It is believed that communication and understanding in an AmI healthcare monitoring information management will be improved.

7 EVALUATION OF THE RESULTS

7.1 Overview

Various factors can affect the quality of a research. Accordingly, as described in the methodology Chapter various evaluation methods were employed to evaluate and validate the work in this Thesis. It is found that both the process and result of this Thesis are valid and reliable. That means given the stated criteria, the research result adequately describes the problem of interest at hand.

This Thesis follows a design science approach and qualitative research and it also complies with its evaluation guideline forwarded in literature. According to, von Alan et al et al., , the utility, quality and efficacy of design artifacts must be rigorously established through well-executed evaluation methods.

It is also mentioned that the selection of metrics such like (completeness, consistency, ease of use, reliability, usability, fit with the organization and purpose, and other relevant attributes) and evaluation methods such as (Surveying, Observational Experimental, Analytical, Descriptive, Testing) is crucial .

Hence, in line with this, the evaluation process in this Thesis agrees with the metrics and methods recommended in literature.

In this Thesis, we followed the assertion made in , as "*we emphasize that evaluation should not be seen as an isolated process activity but needs to be considered from the beginning of a design process.*",

Issues of validity, reliability and generalizability of findings, accuracy and interestingness of machine learning experiments, completeness, practical utility, and robustness of the architectural models are examined and explained in the following Subsections.

7.2 Validity and Reliability of the Process and Results

According to Creswell, Gibbs , in qualitative researches, validity means that the researcher checks for accuracy of findings by employing certain procedures, while reliability refers to the consistency of the approach used. In this Section, the discussion is basically following the validity and reliability issues that were presented in the methodology Chapter. Accordingly, the theoretical support with the quality of the empirics and data collection and analysis is described in this Section.

The Theoretical Support in the Research (Descriptive Evaluation)

An important way of validating a research work is to show the theoretical support. According to Xie and Helfert, von Alan. , a descriptive evaluation method involves informed argument using information from relevant researches to build a convincing argument for the process of building artifacts and their utility. Accordingly, effort has been made to refer relevant literature in the area throughout the whole research process.

The literature review that has been used in this dissertation has been judged by their credibility. By doing this, we believe to improve the validity and reliability of both the process and the end result. Major consideration in assessing the credibility include, slight consideration of publisher, the date of the source and the source matching to the topic.

Regarding the framework employed, **ZF**, as discussed in , has many important features such as its capability, usefulness, simplicity and a problem-solving tool. It is also mentioned that it communicates, helps planning, which proves its wide applicability. Its applicability in security, planning, and the architecture is also demonstrated in , as *"It is shown that Zachman Framework best fits to plan security architecture for an enterprise as any evolving changes in technology can be implemented onto the Zachman Framework without affecting the direction of the enterprise."*

Although, it appears that **ZF** has not been used yet to define **AmIHCMIA**, it has been proved to be useful in areas related to information technology like network security, road safety, digital library content, enterprise aggregate planning. Thus, as it is shown here in, using a **ZF** based architectural descriptions improves communication by reducing misunderstandings and obstacle in defining proper information management practices. The use of **DM** Process model also confirms the valid theoretical base for data mining experiments conducted.

The Empirics in the Research

As this research at hand is a qualitative case study, mainly, interview was used as a major data collection technique, combined by observation and documents. In the same line, the concept of validity in a qualitative approach, effort has been made to present rich description to transfer findings. Accordingly to Creswell , detailed descriptions of the settings and findings make a research output more realistic and richer. Where themes are described from different data sources namely; interview and group discussion, observation and document analysis, was also used. The process of supporting themes based on converging sources of data adds on the validity of research findings .

Another important undertaking in the process of the validation of this research is member-checking methods. In connection to this, efforts have been made to determine the accuracy of the findings through by taking the presentations, understanding and the taking the resulting models back to the research participants for comment and confirmation. It is also important to illuminate the capability of the participants in understanding the research results. Accordingly, the fact that they are experts in the area and working for longer time enables the participants to reasonably understand the concepts and presentations regarding the research findings.

Regarding the reliability of the work and findings, the data collection and analysis procedures described in the methodology and result section provide the necessary details to justify the consistency of the research approach. Another worth mentioning issue in the research is the generalizability of findings and results. Accordingly to Yin , the generalization occurs when researchers study additional cases and generalize findings to new cases. Following this argument, it is easy to understand that the results of this research can be further expanded and generalized by studying different cases.

7.3 Accuracy of the Experimental Results

With respect to the validation of the DM experiments, experimental techniques recommended in literature were employed. Accordingly, objective like accuracy (classification success), error rate (misclassification), ROC, Sensibility were applied.

Accordingly, the best models selected exhibited acceptable accuracy (over 95%) and ROC score with lower error rate as presented in previous Chapters. Moreover, from the accuracy point of view, Random committee model can be identified as the best choice for analysis and detection model among all the other individuals' classifiers and meta classifiers. Random committee provides an advantage that with a reduced feature set a better classification accuracy and performance and is able to offer a better decision support system. However ensemble classifiers through voting technique combined with three or five classifiers provided better accuracy results and are found to be effective in analysis of AmIHCM.

Ensemble combining (Voting + 3) or (Voting + 5) classifiers are very efficient and can achieve high accuracy and, better outcomes that are significantly better compared with the outcomes of the all the base individual classifiers proposed, all meta base classifiers and all ensemble combined methods.

7.4 Completeness, Robustness and Practical Utility of HCMIA (Survey)

This Subsection discusses the procedure and actual evaluation survey results from the respondents.

7.4.1 Evaluation Procedure

A formative evaluation was conducted by the principal researcher. The purpose of the evaluation is to test that the study and design of **AmIHCMIA** is a working idea. For this purpose, the evaluation involved a presentation to show the core idea of **AmI** concepts, and also idea of studying and designing **AmIHCMIA** was functional and feasible. Both the presentation and following evaluation process are conducted at the twenty-five hospitals in Khartoum state by principal researcher as indicated in the methodology chapter.

A survey instrument is also formulated and distributed with a relevant questionnaire items and interview questions about the concepts and the architectural description. To improve the reliability, the survey items are adopted from previous researches and modified to fit the purpose. The survey items are attached as in Appendix F**.** The result of the evaluation process through this survey is presented in the next Sub-section

7.4.2 Evaluation of Survey Results

The evaluation method involved supplying a copy of the architectural descriptions and explanation of **AmI** concept and each concept in the **IA** to the participants, who are actually experts in healthcare vital signs monitoring data collection and analysis. Twenty-five healthcare vital signs monitoring experts whose work is directly related to vital signs monitoring data management have participated in the evaluation process.(see sample of participation answering Appendix G)

As discussed in the previous section, to confirm the validity of the **AmIHCMIA**, 30 questionnaires containing 24 likert type items and 5 open-ended questions were distributed among experts in healthcare vital signs monitoring departments in both private and public hospitals at Khartoum state. Finally, 25 completed questionnaires were returned and used for analysis. The questionnaire's reliability was analyzed using one of the reliability analysis features of SPSS 16.0 tool.

Accordingly, the Cronbach's Alpha ($\alpha = 0.813$) was calculated (see Appendix H) which confirms its reliability. Regarding the analysis of experts' response for Likert type items, standard deviation and mean of descriptive statistics were used. In view of that the standard deviation was calculated for each item of the survey based on the gathered data. As evident from the table in Annex A, it is easy to deduce that the healthcare monitoring IA is acceptable as the survey exhibits standard deviation less than 1 (0.569 – 0.935) for each item.

Another parameter used in the analysis of the experts' response on the acceptability of the IA is comparing the mean score results of completed questionnaire with the questionnaire's average. Hence, the mean score of the 25 completed questionnaires was 83.33, which is by far more than the questionnaires average score, 75. Respondents rate most of the questions in the survey above average, which indicates that the **IA** defined in this research in the opinion of healthcare monitoring experts is desirable.

The next part of the evaluation survey was composed of 5 open-ended questions, for which experts responded accordingly. The first question is about **IA** "What do you understand with the term Information architecture? Does your organization have any?". Accordingly, all the respondents expressed their understanding of the concept **IA** from the discussion made by researcher and they assumed that it is a way of organizing information using information and communication technology.

Therefore, none of the respondent can provide any further explanation about the **IA** concept. In connection to this, it is also noticed that no organized **IA** existed in their respective organization. However, one of the respondents from one of the hospitals claim that they consider their procedure in collecting and analyzing vital signs monitoring information as a form of **IA** but not well guided and systematized which really make sense.

When asked about any advantage or disadvantage of using such type of **IA**, respondents list a number of advantages and a few of disadvantages too. This implies, their significant understanding of the material presented. Major advantages of the use of **IA** includes; facilitate access to and update the monitoring data, create a platform for further investigation, facilitate vital signs monitoring data quality and integration, improve decision support etc..

It was also interesting to learn some possible disadvantages of the use of such type of **IA**. Accordingly, two of the respondents mentioned that use of such type of **IA** requires training, being responsible and cost. However, given the level of priority that the healthcare vital signs monitoring sector has got, seriousness of the problem worldwide and the increasing awareness of the community, we argue that the advantages outweigh the disadvantages.

The third question posed in the evaluation interview is centered on the respondents understanding of the healthcare monitoring **IA** presented. Accordingly, all respondent confirmed that they really understand the presented healthcare monitoring **IA** with a phrase ‘to some extent’ and ‘somehow’ in case of the two respondents.

The fourth question involved if there any aspects of an **IA** missing in opinion of respondents, all respondents confirmed that there were no other aspects missing the proposed **IA**.

The last question was about the expected potential use of the **IA** within their respective organizations. While, two of the respondents have a reservation on the implementation of the defined healthcare monitoring **IA** due to its resource requirement, the other six expected the potential use of it. Some of the reasons mentioned for its potential use include, the government attention to the healthcare monitoring issues and the availability of volunteers in the process of improving healthcare monitoring problems. It is also stated that the architecture will help in reviewing the healthcare monitoring information management practice, which agrees with the potential contribution of this research.

It is worth mentioning that in the proposed architecture, all six columns in **AmIHCMIA** (content, motivation, function, people, place and time) are interchangeable in order, each column has a simple generic model representing one aspect, the basic model of each column is unique, each row describes a distinct perspective, each cell is unique and integration of all cell models in one row constitutes a complete model from the perspective of that row. This is to mean that architecture is in agreement with the integrity rules.

7.5 Summary of the Evaluation Process

Understanding the importance of evaluation of the process and the end results of a given research, a multi-method evaluation technique was employed to evaluate and validate the work. Generally, it is found that both the process and result of this research are valid, reliable and acceptable. This is shown through the theoretical literature support in the process of the research and the experimental and survey evaluation employed to assess the validity and acceptability of the end result.

The reliability of the evaluation survey items itself was also checked using available reliability analysis techniques in addition to the fact that they are adopted from previous researches. In the case of the structured interview the experience and insight that the healthcare monitoring experts provide a confidence on their better position to comment the research results.

Conference and Journal publications as indicated , also support the validity of the process and result of this research undertaking. However, from the very nature of such type of research, it is believed that the ultimate benefit of the **AmIHCMIA** will be seen in the long run as it requires time for its implementation and its impact.

Summary of the evaluation process is presented in Table 7.1. The table summarizes the major aspects of the research evaluated, purpose of the specific evaluation, methods and outcome of each evaluation.

Table 7.1: Summary of evaluation process

	Aspects Evaluated (What)	**Purpose (Why)**	**Methods of evaluation (How)**	**Evaluation Outcome**
1	AmI Healthcare Monitoring Information Architecture	To confirm its completeness, clarity and Relevance To test that the study and design of AmI healthcare monitoring is a working idea	Surveying Descriptive	AmI healthcare monitoring IA is acceptable
2	Wearable Sensors Vital Signs Data Analysis Models	To measure accuracy and interestingness of the models and resulting patterns	Experimental (Accuracy, Error Rate, ROC)	Models perform well and Accuracy is acceptable
3	The research process	To confirm the validity and reliability of the research	Literature support Expert participation	

8 CONCLUSION AND FUTURE RESEARCH

8.1 Overview

In this Chapter, we present conclusions and summary of the results. The following Sections present summary, contributions, limitations of the research and conclusions as results of meeting the objectives of the research. Finally, we also identified future research so as to let others researchers continue in exploring the area and improve the results.

8.2 Summary

In this research, attempt has been made to investigate data mining experiments in explaining health care monitoring patient's situations and investigation of enterprise **IA** concepts. The main goal of this research is to investigate novel ensemble decision support model by intelligent analysis of patient's sensor data using data mining tools, to get better results and to improve assisted health care monitoring.

In doing so, our experiments were conducted on simulated wearable sensors vital signs monitoring data in a hospital environment. Survey of the literature enabled to create understanding of the techniques and attempts in a healthcare wearable sensors vital signs monitoring data quality and data analysis domain.

We reduced the number of attributes from 300 attributes to 6 attributes. We explored various ensembles combination models and evaluated the models with various methods to evaluate performance based on Error Metrics, **ROC** curves, Confusion Matrix, Sensitivity, Specificity and the Cost/Benefit methods.

We compared the performance of the entire classifiers and empirical results illustrate that Voting combined with J48, Random Forest, Random Tree (Voting + 3 classifiers) model; with selection attribute method gives better accuracy, with high recall and high f- measure. Our Novel Intelligent Ensemble Health Care Decision Support and Monitoring can optimize the results and improve assisted health care monitoring.

Another major aspect of this research is the definition of healthcare monitoring **IA**. Extensive survey of literature in the area has been made as discussed in literature review chapter to provide evidence for the applicability of enterprise **IA** in other areas. In line with this, a review of literature also enabled us to create a good understanding of international practices and attempts in improving patients monitoring data collection and analysis in a healthcare monitoring domain.

Accordingly, this research presents **IA** for healthcare monitoring data collection and analysis systems. The defined architecture is based on **ZF** separation of concerns. By applying the top three layers of **ZF** hierarchy, it was possible to develop descriptive **AmIHCMIA** that can facilitate communication. The development of **AmIHCMIA** in this research and design of an **AmI** healthcare monitoring **IA** is an original contribution, which improves and expands the conceptual framework of the research in both healthcare monitoring domain and **IA** field. **AmI** healthcare monitoring **IA** can serve as a strategic guide to the review and development of the healthcare monitoring data collection and analysis systems. It can also be used as a tool for analysis and re-engineering of existing monitoring data systems.

The result of the research helps healthcare monitoring organizations to revisit their focus of attention in drafting and implementing measures to reduce healthcare monitoring problems.

More specifically, the research indicated that in addition to metros, nurses and doctors should address well others departments in the hospital and the relatives

It is also worth mentioning that systematic data collection and quality check along with periodic analysis should get due attention so that other measures will be knowledge driven. The research result can also be used as it is or replicated in other developing countries with similar environment in the area of patient’s vital signs monitoring data collection and analysis.

Regarding evaluation of the research results, various evaluation method including surveying, experiments and descriptive techniques were employed to evaluate and validate the work. Generally, it is found that both the process and result of this research are valid and acceptable. This is shown through the theoretical literature support in the process of the research, the experimental and survey evaluation employed to assess the validity and acceptability of the end result.

8.3 Contributions of the Research

This research has investigated existing theoretical and conceptual frameworks on **IA** and patient's vital signs monitoring analysis, and established the use in the domain of healthcare monitoring information management. It further aimed to broaden these ideas by explaining healthcare monitoring situations and bringing in components identified in related research areas, with the ultimate goal being the development of integrated **AmIHCMIA** for healthcare monitoring domain. In meeting the research objectives, the work has generated a number of outcomes that are contributions to the knowledge and practices.

8.3.1 Contributions to the knowledge

The major contributions of this research are the design of the architectural descriptions collectively named as the **AmIHCMIA** and knowledge embodied in it. As **AmIHCMIA** is the first architectural description to address healthcare monitoring information management from an enterprise perspective, its design itself is a contribution to design science.

Another interesting contribution includes the design process, as it was possible to clearly indicate the process as a base for the design of the architecture. These contributions advance our understanding of how best to structure information assets. The details of these contributions are presented below.

Architectural description of healthcare monitoring information and process, which will be used to evaluate existing systems and/or design a new one, is a major contribution to both the design science research and the **IA**. Detail description of the content, process, motivation, network, people and time provides a comprehensive view in addressing healthcare monitoring issues from information management point of view.

Analyzing the nature of the problem area and introducing enterprise view in a healthcare-monitoring domain is another contribution.

Sensing the absence of overarching architectural guide and the disintegration in the effort of addressing healthcare monitoring problem, this work provides an enterprise view so that stakeholder can view the problem domain from different perspectives. As there was no any healthcare monitoring **IA** from enterprise perspective, this is an important contribution to the **IA** knowledge area.

The advantage of having an **IA** from an enterprise view, specifically for healthcare monitoring organizations includes providing a platform for standardizing the content and process of healthcare monitoring information management. With this regard issues related to patient's vital signs monitoring sources and contents were explored.

It also provides a research framework for future efforts of improving healthcare monitoring. It is also worth mentioning the methodology contribution to the **IA** area on how to develop architectural description under **ZF** in a specific domain.

As discussed in Chapter 6, the design process in pulling data mining results and empirical qualitative data guided by the theoretical support to define **IA** is a contribution.

This is because there is no as such prior research attempting to integrate data mining results and qualitative data using varied modeling tools in developing architectural descriptions. This is inevitable as the result and processes of this Thesis are found to be acceptable through the evaluation process.

8.3.2 Contribution to the Practice and Recommendations

Accordingly, based on the architectural model artifacts, which are results of the review of international practices, successive **DM** experiments, empirical data from healthcare monitoring departments and study of relevant literature, the following basic implications and recommendations to the practice are identified.

Explanation of the healthcare monitoring situation is one of the major aspects of the research at hand. This was possible through experimenting and suggesting analytical machine learning models in describing the nature and magnitude of the healthcare monitoring problem.

Classifications models in determining the important attributes contributing to patient's vital signs occurrence and severity add on the effort being made in understanding healthcare monitoring situations. Hence, selected patient's vital signs monitoring analysis models can be integrated to make periodic analysis of patient's vital signs monitoring data, which is believed to improve measures, which are an important contribution to the healthcare monitoring domain. In this regard the use of multi-classifier systems and the investigation of novel ensemble combine decision support model is one of the main aspects in this research contributing to both the healthcare monitoring and the analysis of healthcare monitoring data management. The data quality issues and patient's vital signs monitoring data analysis trend conducted through this research were also worth mentioning contributions.

Investigation of **ZF**, in a **AmIHCM** domain is also another aspect to provide a framework for the integration of future standards developed for data representation, manipulation, and visualization. **ZF** has been used in areas like network security planning, road safety, education services delivery, determining the content of digital libraries.

This work extends the use of the framework in healthcare monitoring domain for a **AmIHCM** information management, which will improve understanding and facilitate communication among healthcare monitoring organizations.

With respect to data handling and information exchange, as evident from international practices and due to structural issues in the study area, the master data source is recommended to be at the healthcare monitoring Manager department.

However, healthcare monitoring departments, nurses, relatives, and doctors should also have a copy of this healthcare monitoring data excluding privacy related attributes (see Figure 6.2). This will allow the healthcare monitoring manager department at various levels to conduct exploratory classification, analysis and decisions as shown in previous experiments , which will be used to continually revise and update. patient's vital signs monitoring data analysis and dissemination of its result is another very important function of a healthcare monitoring system in patient's vital signs monitoring information management.

It is also recommended to conduct classification and exploratory experiments by integrating models into a system. The architecture therefore will allow the integration of open-source analysis tools in to the overall system. Therefore, it is anticipated that with improvements in the data content, collection and reporting, patient's vital signs monitoring analysis can better be enhanced.

Though the research work has resulted in a number of contributions as mentioned above, the completion of this work does not denote the end of the researcher's study of **AmIHCM** and **IA**. It rather marks the beginning by creating an empirical, theoretical and wide-ranging basis for establishing best way to structure information and successful AmI healthcare monitoring information management. It is believed that this research can have a significant impact on the state of patient's vital signs monitoring information management as healthcare monitoring organizations are convinced to make use of the **IA**.

8.4 Limitations of the Research

This research has sought to use a multifaceted approach: a qualitative empirical data using interview, observation and document analysis; experiments using various data mining techniques; critical literature review and researchers past expertise. Despite the various efforts to overcome risks on the quality of the research findings and the process employed, some limitations must be recognized.

One such possible limitation of this research undertaking is on the validation of the architectural description. In relation to this, attempt has been made to provide thick description on the empirical data collection and analysis as well as base on multiple references in the development of architecture in the research process. As to the evaluation process, use of questioning approach through survey items and employing descriptive evaluation through theoretical support were the major ones.

Thus, particularly the questioning approach through survey items might not be sufficient to practically show the usefulness of the architectural description as it only investigates the healthcare monitoring (domain) expert's opinion on the healthcare monitoring **IA**. The inclusion of information technology experts would have made the evaluation stronger.

In addition, in order for comprehensive evaluate to the architecture in a real environment, it would be necessary to demonstrate through developing (prototype) and have healthcare monitoring organizations adopt this enterprise view based **AmIHCMIA** and then evaluate if the process of patient's vital signs monitoring information management is improved and justify the impact on the improvement of healthcare monitoring situations in general.

However, this would require a longitudinal research, which needs a longer time and is beyond the scope of this research. The performance of the **DM** experiments is better compared to the previous experiments done in the research domain area.

The models were built based on training data (which is based on the number of patients at a time) and the performance might change for another set of data. It would be required to re-train the models if the number of patients changes. However the training time is very fast. Scalability is another research issue. It would be interesting to know how such a model would work in real life when 1000's of patients need to be monitored at the same time.

8.5 Future Research Direction

In this research novel ensemble combine decision support method by intelligent analysis of simulated patient's sensor data and integrated with **AmIHCMIA** are developed and through the process it is established that future research could examine issues in AmI healthcare monitoring information management in an integrated manner. The result of this research can be used to support future research related to AmI healthcare monitoring and application of **IA** concepts, especially in the context of healthcare monitoring in general and **AmIHCM** in particular. Hence, using the same ensemble combine method and integrated **IA** approach as defined for the AmI healthcare monitoring information management, other information requirements can also be addressed and provide organizations with a more effective way of managing information assets.

Accordingly, the research project will carry on, in collaboration with the relevant healthcare monitoring organs, by using the achievements of this research as a guide to evaluate the existing patient's wearable sensors vital signs monitoring information management and documenting the status in others states in Sudan. The aim of this additional research project could be either to evaluate the AmI healthcare monitoring as analyzing instrument or to evaluate it as an information system development tool. This will be followed by the implementations of achievements of this research to the identified problems. One of the major outputs of this research, **AmIHCMIA**, assumes that the target be the patient's wearable sensors vital signs monitoring data collection and analysis at this stage.

Future research direction could also be on establishing an architecture framework for integration with others hospital healthcare systems. Moreover, based on the major results of this research, it is logical to recommend that further investigation is required to expand and enrich the defined architecture. A more detail and diversified classification analysis methods of patient's vital signs monitoring using a range of as well as a combination of **DM** techniques is also another potential area of future work. In connection to this, the result of this study can also be used to support future research related to **DM** approaches such as ensemble technique in AmI in general and especially in the context of AmI healthcare monitoring.

8.6 Conclusion

In order for healthcare monitoring organizations achieve their objective of addressing healthcare monitoring problems; they must be excel in **AmIHCM** information management. The current approach to patient's vital signs monitoring information management is not providing analysis to patient's vital signs monitoring data. Also the current approach is not providing the various stakeholders with the information and report; they require in order making effective decisions. This is due to the fact that healthcare monitoring organizations departments at various levels are working in a fragmented manner and without analysis approaches.

The solution to the problem is to develop a model based on patient's data analysis and the usage of **IA**. By using **IA** approach, healthcare monitoring organizations develop a better understanding of the content, motivation, process, place, time and people of patient's vital signs monitoring information management.

To achieve the main objective and sub-objectives of this research:. Firstly various concepts related to **AmIHCM**, **DM** approaches, **IA** and the guiding framework are described; Second, various data mining experiments, ensemble model was developed, qualitative data collection and analysis were done to determine information requirements. Finally the proposed **AmIHCMIA** based on **ZF** was constructed.

Thus, this dissertation resulted in a suitable novel ensemble assisted healthcare monitoring model also suitable enterprise **IA** model, which has been related to relevant **IA** goals and information sources. It is believed that as the practice evolves through time and gets more advancement through subsequent research works, it will be empirically proven in real world use.

References

Appendices

Appendix A - Experimental results of the performance of different classification techniques

10 Selection Attribute		6 Selection Attribute	
Test Options: Use training set.		**Test Options: Use training set.**	
Classifer	**Correctly Classified**	**Classifer**	**Correctly Classified**
BayesNet	60.2685 %	BayesNet	80.8054 %
NaiveBayes	56.3758 %	NaiveBayes	73.9597 %
NaiveBayes Multinomial	60.5369 %	NaiveBayesMultinomial Text	52.8859 %
NaiveBayes MultinomialText	52.8859 %	NaiveBayesUpdateable	73.9597 %
NaiveBayes MultinomialUpdateable	60.5369 %		
NaiveBayesUpdateable	56.3758 %		
Functions			
Logistic	63.8926 %	Logistic	75.0336 %
MultilayerPerceptron	49.2617 %	MultilayerPerceptron	88.1879 %
SGD	61.3423 %	SGD	75.5705 %
SGDText	52.8859 %	SGDText	52.8859 %
SimpleLogistic	61.3423 %	SimpleLogistic	75.0336 %
SMO	51.0067 %	SMO	74.094 %
VotedPerceptron	52.8859 %	VotedPerceptron	54.2282 %
Lazy			
IBk	63.8926 %	IBk	100 %
KStar	63.3557 %	KStar	95.5705 %
LWL	60.2685 %	LWL	84.8322 %
meta			
AdaBoostM1	60.4027 %	AdaBoostM1	79.3289 %
AttributeSelectedClassifier	61.4765 %	AttributeSelectedClassifier	97.9866 %
Bagging	63.8926	Bagging	95.9732 %
ClassificationViaRegression	61.3423 %	ClassificationViaRegression	93.2886 %
CVParameterSelection	52.8859 %	CVParameterSelection	52.8859 %
FilteredClassifier	60.2685 %	FilteredClassifier	83.2215 %
LogitBoost	62.953 %	LogitBoost	85.906 %

MultiClassClassifier	63.8926 %	MultiClassClassifier	75.0336 %
MultiClassClassifierUpdateable	61.3423 %	MultiClassClassifierUpdateable	75.5705 %
MultiScheme	52.8859 %	MultiScheme	52.8859 %
RandomCommittee	63.8926 %	RandomCommittee	100 %
RandomSubSpace	63.8926 %	RandomSubSpace	93.2886 %
Stacking	52.8859 %	Stacking	52.8859 %
Vote	52.8859 %	Vote	52.8859 %
Misc			
InputMappedClassifier	52.8859 %	InputMappedClassifier	52.8859 %
Rules			
DecisionTable	62.0134 %	DecisionTable	89.7987 %
JRip	62.8188 %	JRip	93.6913 %
OneR	63.8926 %	OneR	80.2685 %
PART	63.6242 %	PART	97.9866 %
ZeroR	52.8859 %	ZeroR	52.8859 %
Trees			
DecisionStump	57.1812 %	DecisionStump	71.1409 %
J48	63.0872 %	J48	97.9866 %
LMT	63.8926 %	LMT	97.5839 %
RandomForest	63.8926 %	RandomForest	99.8658 %
RandomTree	63.8926 %	RandomTree	100 %
REPTree	63.8926 %	REPTree	90.7383 %
Test Options: Cross-Validation		**Test Options: Cross-Validation**	
Bayes			
BayesNet	59.5973 %	BayesNet	78.7919 %
NaiveBayes	51.6779 %	NaiveBayes	73.8255 %
NaiveBayesMultinomial	60.5369 %		
NaiveBayesMultinomialText	52.8859 %	NaiveBayesMultinomialText	52.8859 %
NaiveBayesMultinomialUpdateable	60.5369 %		
NaiveBayesUpdateable	51.6779 %	NaiveBayesUpdateable	73.8255 %
Functions			
Logistic	63.6242 %	Logistic	74.3624 %
MultilayerPerceptron	57.1812 %	MultilayerPerceptron	84.2953 %
SGD	54.8993 %	SGD	73.9597 %
SGDText	52.8859 %	SGDText	52.8859 %
SimpleLogistic	57.4497 %	SimpleLogistic	73.6913 %
SMO	55.0336 %	SMO	73.1544 %

VotedPerceptron	52.8859 %	VotedPerceptron	53.0201 %
Lazy			
IBk	63.8926 %	IBk	90.3356 %
KStar	62.0134 %	KStar	89.5302 %
LWL	59.8658 %	LWL	80.9396 %
Meta			
AdaBoostM1	58.3893 %	AdaBoostM1	80.4027 %
AttributeSelectedClassifier	60.6711 %	AttributeSelectedClassifi er	91.9463 %
Bagging	62.2819 %	Bagging	90.4698 %
ClassificationViaRegressin	57.3154 %	ClassificationViaRegress in	88.4564 %
CVParameterSelection	52.8859 %	CVParameterSelection	52.8859 %
FilteredClassifier	57.7181 %	FilteredClassifier	85.906 %
LogitBoost	61.0738 %	LogitBoost	82.2819 %
MultiClassClassifier	63.6242 %	MultiClassClassifier	74.3624 %
MultiClassClassifierUpdateabl e	54.8993 %	MultiClassClassifierUpd ateable	73.9597 %
MultiScheme	52.8859 %	MultiScheme	52.8859 %
RandomCommittee	63.8926 %	RandomCommittee	95.0336 %
RandomSubSpace	61.8792 %	RandomSubSpace	88.5906 %
Stacking	52.8859 %	Stacking	52.8859 %
Vote	52.8859 %	Vote	52.8859 %
Misc			
InputMappedClassifier	52.8859 %	InputMappedClassifier	52.8859 %
Rules			
DecisionTable	60.1342 %	DecisionTable	84.8322 %
JRip	60.9396 %	JRip	89.5302 %
OneR	62.0134 %	OneR	73.557 %
PART	62.1477 %	PART	92.2148 %
ZeroR	52.8859 %	ZeroR	52.8859 %
Trees			
DecisionStump	57.1812 %	DecisionStump	67.2483 %
J48	62.0134 %	J48	92.8859 %
LMT	62.6846 %	LMT	92.2148 %
RandomForest	62.4161 %	RandomForest	94.2282 %
RandomTree	63.8926 %	RandomTree	94.8993 %
REPTree	61.745 %	REPTree	88.1879 %
Test Options: Percentage split 66%		**Test Options: Percentage split 66%**	

bayes			
BayesNet	59.2885 %	BayesNet	76.2846 %
NaiveBayes	56.917 %	NaiveBayes	73.913 %
NaiveBayesMultinomial	60.4743 %		
NaiveBayesMultinomialText	56.917 %	NaiveBayesMultinomial Text	56.917 %
NaiveBayesMultinomialUpdat eable	60.4743 %		
NaiveBayesUpdateable	56.917 %	NaiveBayesUpdateable	73.913 %
Functions			
Logistic	62.0553 %	Logistic	74.3083 %
MultilayerPerceptron	52.5692 %	MultilayerPerceptron	84.9802 %
SGD	60.0791 %	SGD	73.5178 %
SGDText	56.917 %	SGDText	56.917 %
SimpleLogistic	60.0791 %	SimpleLogistic	73.5178 %
SMO	53.7549 %	SMO	71.5415 %
VotedPerceptron	43.083 %	VotedPerceptron	56.5217 %
Lazy			
IBk	62.0553 %	IBk	88.9328 %
KStar	61.2648 %	KStar	88.9328 %
LWL	62.4506 %	LWL	83.3992 %
meta		**meta**	
AdaBoostM1	62.4506 %	AdaBoostM1	80.6324 %
AttributeSelectedClassifier	60.0791 %	AttributeSelectedClassif ier	89.3281 %
Bagging	54.9407 %	Bagging	88.1423 %
ClassificationViaRegression	60.8696 %	ClassificationViaRegres sion	88.5375 %
CVParameterSelection	56.917 %	CVParameterSelection	56.917 %
FilteredClassifier	59.2885 %	FilteredClassifier	84.585 %
LogitBoost	61.6601 %	LogitBoost	83.7945 %
MultiClassClassifier	62.0553 %	MultiClassClassifier	74.3083 %
MultiClassClassifierUpdateabl e	60.0791 %	MultiClassClassifierUpd ateable	73.5178 %
MultiScheme	56.917 %	MultiScheme	56.917 %
RandomCommittee	62.0553 %	RandomCommittee	92.8854 %
RandomSubSpace	62.0553 %	RandomSubSpace	84.1897 %
Stacking	56.917 %	Stacking	56.917 %
Vote	56.917 %	Vote	56.917 %
misc			
InputMappedClassifier	56.917 %	InputMappedClassifier	56.917 %

rules			
DecisionTable	60.0791 %	DecisionTable	84.9802 %
JRip	59.6838 %	JRip	86.166 %
OneR	61.2648 %	OneR	69.17 %
PART	62.4506 %	PART	92.4901 %
ZeroR	56.917 %	ZeroR	56.917 %
trees			
DecisionStump	59.2885 %	DecisionStump	69.17 %
J48	61.6601 %	J48	89.3281 %
LMT	62.0553 %	LMT	89.7233 %
RandomForest	61.2648 %	RandomForest	94.0711 %
RandomTree	62.0553 %	RandomTree	92.0949 %
REPTree	62.4506 %	REPTree	88.9328 %

Appendix B: Results of Met Classifiers in term of TP Rate and FP Rate

Meta	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
AdaBoostM1	0.755	0.152	0.815	0.755	0.784	0.607	0.874	0.859	Normal
	0.848	0.245	0.795	0.848	0.821	0.607	0.874	0.868	Abnormal
AttributeSelectedClassifier	0.915	0.076	0.915	0.915	0.915	0.838	0.953	0.933	Normal
	0.924	0.085	0.924	0.924	0.924	0.838	0.953	0.947	Abnormal
Bagging	0.906	0.096	0.893	0.906	0.900	0.809	0.966	0.964	Normal
	0.904	0.904	0.915	0.904	0.909	0.809	0.966	0.961	Abnormal
ClassificationViaRegressin	0.886	0.117	0.871	0.886	0.879	0.769	0.949	0.950	Normal
	0.883	0.114	0.897	0.883	0.890	0.769	0.949	0.946	Abnormal
CVParameterSelection	0.000	0.000	0.000	0.000	0.000	0.000	0.496	0.469	Normal
	1.000	1.000	0.529	1.000	0.692	0.000	0.496	0.526	Abnormal
FilteredClassifier	0.875	0.155	0.834	0.875	0.854	0.719	0.919	0.908	Normal
	0.845	0.125	0.883	0.845	0.864	0.719	0.919	0.914	Abnormal
LogitBoost	0.801	0.157	0.819	0.801	0.810	0.644	0.893	0.883	Normal
	0.843	0.199	0.826	0.843	0.834	0.644	0.893	0.877	Abnormal
MultiClassClassifier	0.732	0.246	0.726	0.732	0.729	0.486	0.841	0.819	Normal
	0.754	0.268	0.760	0.754	0.757	0.486	0.841	0.847	Abnormal

MultiClassClassifierUpdateable	0.741	0.261	0.716	0.741	0.728	0.479	0.740	0.653	Normal
	0.739	0.259	0.762	0.739	0.750	0.479	0.740	0.701	Abnormal
MultiScheme	0.000	0.000	0.000	0.000	0.000	0.000	0.496	0.469	Normal
	1.000	1.000	0.529	1.000	0.692	0.000	0.496	0.526	Abnormal
RandomCommittee	0.957	0.056	0.939	0.957	0.948	0.901	0.984	0.974	Normal
	0.944	0.043	0.961	0.944	0.953	0.901	0.984	0.983	Abnormal
RandomSubSpace	0.829	0.063	0.921	0.829	0.873	0.873	0.955	0.933	Normal
	0.937	0.171	0.860	0.937	0.897	0.773	0.955	0.966	Abnormal
Stacking	0.000	0.000	0.000	0.000	0.000	0.000	0.496	0.469	Normal
	1.000	1.000	0.529	1.000	0.692	0.000	0.496	0.526	Abnormal
Vote	0.000	0.000	0.000	0.000	0.000	0.000	0.496	0.469	Normal
	1.000	1.000	0.529	1.000	0.692	0.000	0.496	0.526	Abnormal

Appendix C: Interview questions

1. How do you explain the healthcare monitoring situation in Khartoum State?
2. How do you describe the effectiveness of healthcare monitoring data collection and analysis practice?
3. What are the available healthcare monitoring data sources?
4. What aspects of healthcare monitoring should be recorded?
5. Do you believe that the healthcare monitoring reporting format being used is complete enough in recording all details about patients?
6. What sort of information you think is missing in the healthcare monitoring reporting ?
7. Can you describe please healthcare monitoring data reporting and analysis process in general?
8. Do you feel that the analysis level is sufficient for the purpose required by the users?
9. How do you deal with healthcare monitoring data?
10.What sort/type of analysis do you make on the healthcare monitoring data at your organization?
11. Are statistical analysis/ machine learning methods involved in your daily work?
12. What aspect of healthcare monitoring information and prediction models should get priority in Khartoum context?
13. How is healthcare monitoring data updated?
14. What is the prime motivation of managing healthcare monitoring data and information?
15. Who are the primary users of healthcare monitoring information?
16. What sort of information is required by stakeholders regarding healthcare monitoring?
17. Who do you think is responsible in healthcare monitoring reporting?
18. Who do you think should participate in healthcare monitoring data management and analysis?
19. Can you mention a specific place (like organizational units) where data reporting, analysis and dissemination should happen?
20. Would you comment please on the timing for healthcare monitoring data reporting, analysis and dissemination?

Appendix D: Interview questions in Arabic language

جامعة السودان للعلوم والتكنولوجيا

كلية الدراسات العليا - كلية علوم الحاسب وتقنية المعلومات

برنامج الدكتوراه عن طريق المقررات الدراسية والأطروحة

المقابلات الشخصية

جمع البيانات الخاصة بمعمارية المعلومات لرصد الرعاية الصحية

فى مستشفيات ولائية الخرطوم

سيتم استخدام البيانات التي سوف يتم جمعها من هذا الاستبيان لتحليل البيانات كجزء من رسالة الدكتوراة للتحقق من معمارية المعلومات فى مجال رصد الرعالية الصحية والتحديات التي تواجه تطويرها في المستشفيات العامة و الخاصة فى ولاية الخرطوم.

سيتم التعامل مع الإجابات التى سوف يتم جمعها من هذا الأستبيان و المقابلات الشخصية بسرية تامة و من غير ذكر الاسماء و لأغراض البحث العلمى فقط.

معلومات المشارك

الأسم :..
..

الوظيفة :...
..

أسم المؤسسة:...

مكان المؤسسة :...

أسئلة المقابلة

1. كيف يمكن تفسير وضع المراقبة الرعاية الصحية في ولاية الخرطوم؟

كيف تصف فعالية جمع بيانات رصد الرعاية الصحية و**عمليات** التحليل؟ .2

ما هي مصادر بيانات رصد الرعاية الصحية المتاحة؟ .3

ما هي جوانب رصد الرعاية الصحية التى يجب أن تسجل؟ .4

هل تعتقد أن شكل تقارير رصد الرعاية الصحية المستخدمة كاملة بما فيه الكفاية في تسجيل .5 جميع التفاصيل عن المرضى؟

أي نوع من المعلومات التي تعتقد إنها مفقود في تقارير رصد الرعاية الصحية؟ .6

هل لك أن تصف تقارير بيانات رصد الرعاية الصحية و عملية التحليل بشكل عام؟ .7

هل تشعر أن مستوى التحليل هو كاف للغرض المطلوب من قبل المستخدمين؟ .8

كيف تتعامل مع بيانات رصد الرعاية الصحية؟ .9

أي نوع من التحليل يتم إستخدامه على بيانات رصد الرعاية الصحية في مؤسستك؟ .10

هل أساليب التحليل الأحصائية وطرق تعليم الآلة تدخل في عملك اليومي؟ .11

ما هى جوانب المعلومات ونماذج التنبؤ لرصد الرعاية الصحية يجب ان تحصل على الأولوية .12 في إطار ولاية الخرطوم؟

كيف يتم تحديث بيانات رصد الرعاية الصحية؟ .13

ما هو الدافع الرئيسي لإدارة البيانات والمعلومات رصد الرعاية الصحية ؟ .14

من هم المستخدمين الرئيسين لمعلومات رصد الرعاية الصحية؟ .15

أي نوع من المعلومات مطلوبه من قبل الجهات المعنية فيما يتعلق برصد الرعاية الصحية؟ .16

من تعتقد أنه المسؤول عن تقارير رصد الرعاية الصحية؟ .17

من برأيك ينبغي أن يشارك في إدارة و تحليل بيانات رصد الرعاية الصحية ؟ .18

هل لك أن تذكر مكان معين (القسم) الذى يتم إعداد تقارير رصد الرعاية الصحية وتحليلها ،19 وتوزيعها ؟

الرجاء التعليق على توقيت إعداد تقاريربيانات رصد الرعاية الصحية وتحليلها وتوزيعها؟ .20

Appendix E: Sample Data Reduction and Analysis Table

Category	Questions	Key words /concepts	Supporting text	Remark
General	How do you explain the Healthcare monitoring situation in Khartoum state?	Problematic Top priority		Observed
	How do describe the effectiveness of health care monitoring data collection and analysis practice?	Not effective	One respondent commented that: since healthcare monitoring department mainly focuses on the responsibility of patient's vital signs monitoring, the analysis process is absent.	The observation also confirmed the same that healthcare monitoring departments focus only on patient’s vital signs monitoring
Content / What	What are the available Patient’s vital signs monitoring data sources?	Patient’s vital signs sheet form	incompleteness is very serious problem at a hospital level”	Patient’s vital signs sheet form containing only the patients vital signs measurement, patient name, date and time

	What aspects of an healthcare monitoring should be recorded?	Nature of an vital signs, patient name,	*"whenever possible including address, mobil number.and doctor data and contact, relatives, environment".* ".	
	Do you believe that the Healthcare monitoring reporting format being used is complete enough in recording all details about an healthcare monitoring ? If not what sort of information you think is missing?	Underreporting Problems on integrity completeness and usefulness of healthcare monitoring information	*"There are many healthcare monitoring expert Menshied the need of integrity between healthcare monitoring department , doctors ,relatives and others healthcare department in the hospital level*	

Appendix F: Ambient Intelligence healthcare monitoring Information Architecture (AmIHCMIA) – Evaluation Instruments

This Appendix contains a copy of questionnaire items and structured interview

Questions to evaluate the information architecture.

Part I- Questionnaire items

Sudan University of Science and Technology

College of Computer Science and Information

PhD Research Program

The information gathered from this questionnaire will be used for analysis in a research report Investigating of Enterprise Architecture and the challenges to its evolution in Khartoum State Private and Public Healthcare sector.

The results of the in-depth interviews will be used for research purposes only. **The responses to this questionnaire will be treated as strictly confidential.**

	Strongly Agree	Agree	No Comment	Disagree	Strongly Disagree
The description of Personal healthcare monitoring impact, and care monitoring business case (Scope Contextual)- Motivation is clear.					
The description of the scope/context – structure/data is clear					
The description of the essential healthcare monitoring organizations and their functions. – People is clear					
The description of the Important health care monitoring services. – function/process is clear					
The Identification of significant health care monitoring events. – time is clear					
The Identification and description of organization and individual locations. – place is clear					

The description of the Personal healthcare monitoring benefit and objectives. Conceptual - motivation is clear					
The description of the conceptual– structure/data is clear					
The description of the Healthcare monitoring information system workflow. conceptual – People is clear					
The description of the Conceptual activity model of health care monitoring. conceptual -function/process is clear					
The description of the Sequence and timelines of healthcare monitoring. conceptual– time is clear					
The description of the Structure and interrelationship of health care monitoring facilities. conceptual– place is clear					
The description of the System functional requirements.-motivation is clear					
The description of the system– structure/data is clear					
The description of the Health care monitoring information - system human system interface architecture .system– People is clear					
The description of the Application architecture with function and user views. system– function/process is clear					
The description of the Health care monitoring event phases and process components. system– time is clear					
The description of the distributed system architecture. system– place is clear					
The applicability of this architectural model is easy					
The architectural model is comprehensive					
The architectural model is flexible					
The architectural model is appropriate					
The architectural model is relevant					
The architectural model is complete for the scope					

Research Questionnaire in Arabic language

جامعة السودان للعلوم والتكنولوجيا

كلية الدراسات العليا - كلية علوم الحاسب وتقنية المعلومات

برنامج الدكتوراه عن طريق المقررات الدراسية والأطروحة

الأستبيان

جمع البيانات الخاصة بمعمارية المعلومات لرصد الرعاية الصحية

فى مستشفيات ولاية الخرطوم

سيتم استخدام البيانات التي سوف يتم جمعها من هذا الاستبيان لتحليل البيانات كجزء من رسالة الدكتوراة للتحقق من معمارية المعلومات فى مجال رصد الرعالية الصحية والتحديات التي تواجه تطويرها في المستشفيات العامة و الخاصة فى ولاية الخرطوم.

سيتم التعامل مع الإجابات التى سوف يتم جمعها من هذا الأستبيان و المقابلات الشخصية بسرية تامة و من غير ذكر الاسماء و لأغراض البحث العلمى فقط.

معلومات المشارك

الرقم :...
.

الوظيفة :..
..

أسم المؤسسة:...

مكان المؤسسة ....

	أوافق بشدة	أوافق	لا تعليق	لاأوافق	لاأوافق بشدة

وصف أثر رصد الرعاية الصحية الشخصية، و مراقبة الرعاية الصحية داخل المؤسسة واضح. نطاق.					
وصف المجال / إطار و هيكل البيانات واضح.					
وصف أقسام رصد الرعاية الصحية الأساسية ووظائفها واضح.					
وصف خدمات رصد الرعاية الصحية الهامة واضحه. .					
تحديد الأحداث الهامة والتوقيت لرصد الرعاية الصحية واضحة.					
تحديد ووصف الاقسام والمواقع الفردية واضح.					
وصف فائدة رصد الرعاية الصحية الشخصية والأهداف واضحه.					
وصف هيكل البيانات واضح.					
وصف سير عمل نظام المعلومات الصحية للرصد الرعاية الصحية. واضح.					
وصف النموذج المفاهيمي لنشاط رصد الرعاية الصحية واضح.					
وصف التسلسل والجدول الزمني للرصد الرعاية الصحية واضح.					
وصف الهيكلية والعلاقة المتبادلة بين مرافق رصد الرعاية الصحية واضح.					
وصف المتطلبات الوظيفية النظام واضح.					
وصف بيانات النظام واضح.					
وصف معلومات رصد الرعاية الصحية - النظام البشري و نظام واجهة بنية النظام واضح					
وصف بنية التطبيق مع وظائف النظام و واجهات المستخدم واضحة.					
وصف مراحل الحدث لرصد الرعاية الصحية ومكونات العملية واضح.					
وصف معمارية النظام الموزعة واضحة.					
قابلية تطبيق هذا النموذج المعماري هو سهله.					
النموذج المعماري شامل.					

					.النموذج المعماري مرن
					.النموذج المعماري مناسب
					.النموذج المعماري هو ذات صلة بالمجال
					.النموذج المعماري كامل للمجال

Part II- Structured interview questions

Q1. What do you understand with the term Information architecture? Does your organization have any?

Q.2 After viewing the information architecture do you think there are any advantages/ disadvantages to developing and using information architecture to help in the deployment and development of healthcare monitoring and analysis systems?

Q3 Do you understand the AmI healthcare monitoring information architecture presented?

Q4 In your opinion is there any aspects of an information architecture missing?

Q5 Do you expected any use or potential use for information architecture with in your organization? If so what are they?

Appendix G: Sample of Research Questionnaire Participation Answering in Arabic language

بسم الله الرحمن الرحيم

جامعة السودان للعلوم والتكنولوجيا

كلية الدراسات العليا - كلية علوم الحاسب وتقنية المعلومات

برنامج الدكتوراه عن طريق المقررات الدراسية والأطروحة

الأستبيان

جمع البيانات الخاصة بمعمارية المعلومات لرصد الرعاية الصحية

فى مستشفيات ولاية الخرطوم

سيتم أستخدام البيانات التي سوف يتم جمعها من هذا الاستبيان لتحليل البيانات كجزء من رسالة الدكتوراة للتحقق من معمارية المعلومات في مجال رصد الرعالية الصحية والتحديات التي تواجه تطويرها في المستشفيات العامة و الخاصة في ولاية الخرطوم.

سيتم التعامل مع الإجابات التى سوف يتم جمعها من هذا الأستبيان و المقابلات الشخصية بسرية تامة و من غير ذكر الاسماء و لأغراض البحث العلمى فقط.

معلومات المشارك

الرقم :..

الوظيفة :....ممرض..

أسم المؤسسة:..........مستشفى الخرطوم التعليمي..........

مكان المؤسسة ....الخرطوم.................................

	أوافق بشدة	أوافق	لا تعليق	لاأوافق	لاأوافق بشدة
وصف أثر رصد الرعاية الصحية الشخصية، و مراقبة الرعاية الصحية داخل المؤسسة واضح. نطاق.		✓			
وصف المجال / إطار و هيكل البيانات واضح.		✓			
وصف أقسام رصد الرعاية الصحية الأساسية ووظائفها واضح.	✓				
وصف خدمات رصد الرعاية الصحية الهامة واضحه. .		✓			
تحديد الأحداث الهامة والتوقيت لرصد الرعاية الصحية واضحة.		✓			
تحديد ووصف الاقسام والمواقع الفردية واضح.		✓			
وصف فائدة رصد الرعاية الصحية الشخصية والأهداف واضحه.	✓				
وصف هيكل البيانات واضح.		✓			
وصف سير عمل نظام المعلومات الصحية للرصد الرعاية الصحية. واضح.	✓				
وصف النموذج المفاهيمي لنشاط رصد الرعاية الصحية واضح.		✓			

وصف التسلسل والجدول الزمني للرصد الرعاية الصحية واضح.		✓			
وصف الهيكلية والعلاقة المتبادلة بين مرافق رصد الرعاية الصحية واضح.	✓				
وصف المتطلبات الوظيفية النظام واضح.		✓			
وصف بيانات النظام واضح.		✓			
وصف معلومات رصد الرعاية الصحية - النظام البشري و نظام واجهة بنية النظام واضح	✓				
وصف بنية التطبيق مع وظائف النظام و واجهات المستخدم واضحة.		✓			
وصف مراحل الحدث لرصد الرعاية الصحية ومكونات العملية واضح.			✓		
وصف معمارية النظام الموزعة واضحة.					✓
قابلية تطبيق هذا النموذج المعماري هو سهله.		✓			
النموذج المعماري شامل.			✓		
النموذج المعماري مرن.		✓			
النموذج المعماري مناسب.		✓			
النموذج المعماري هو ذات صلة بالمجال.		✓			
النموذج المعماري كامل للمجال.				✓	

Appendix H: Evaluation Result (Descriptive and Reliability)

This Appendix contains evaluation results and its reliability (descriptive and reliability)

DESCRIPTIVES VARIABLES=Qus1 Qus2 Qus3 Qus4 Qus5 Qus6 Qus7 Qus8 Qus9 Qus10 Qus11 Qus12 Qus13 Qus14 Qus15 Qus16 Qus17 Qus18 Qus19 Qus20 Qus21 Qus22 Qus23 Qus24

/STATISTICS=MEAN STDDEV.

Descriptive

[DataSet1] C:\Users\AFAF-PC\Desktop\evaluation result.sav

Descriptive Statistics

	N	Mean	Std. Deviation
Qus1	25	4.12	.881
Qus2	25	4.48	.653
Qus3	25	4.36	.638
Qus4	25	4.08	.702
Qus5	25	4.12	.833
Qus6	25	4.12	.726
Qus7	25	4.44	.583
Qus8	25	4.12	.881
Qus9	25	4.64	.569
Qus10	25	4.12	.881
Qus11	25	4.08	.862
Qus12	25	4.20	.764
Qus13	25	4.24	.597
Qus14	25	4.04	.935
Qus15	25	4.08	.862
Qus16	25	4.68	.748
Qus17	25	4.04	.935
Qus18	25	4.52	.872
Qus19	25	4.28	.737

RELIABILITY

/VARIABLES=Qus1 Qus2 Qus3 Qus4 Qus5 Qus6 Qus7 Qus8 Qus9 Qus10 Qus11 Qus12

Qus13 Qus14 Qus15 Qus16 Qus17 Qus18 Qus19 Qus20 Qus21 Qu s22 Qus23 Qus24

/SCALE('ALL VARIABLES') ALL

/MODEL=ALPHA.

Reliability

[DataSet1] C:\Users\AFAF-PC\Desktop\evaluation result.sav

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	25	100.0
	Excluded[a]	0	.0
	Total	25	100.0

a.Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items

Case Processing Summary

		N	%
Cases	Valid	25	100.0
	Excluded[a]	0	.0
	Total	25	100.0

a.Listwise deletion based on all variables in the procedure.

Reliability Statistics

.813	24

Appendix I: Sample data

A1 fx A

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
2	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
3	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
4	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
5	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
6	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
7	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
8	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
9	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
10	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
11	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
12	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
13	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
14	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
15	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
16	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
17	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
18	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
19	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
20	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
21	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
22	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
23	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
24	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0
25	2935	4617	781	897	39	4267	4360	28	385	387	68	615	9364	49	49	4127	615	96	4685.77	1166700	0